



T.C.

SELÇUK ÜNİVERSİTESİ

FEN BİLİMLERİ ENSTİTÜSÜ

VERİ MADENCİLİĞİ SÜREÇLERİ İLE AKDENİZ BÖLGESİ GÜNEŞ IŞINIM POTANSİYELİNİN TAHMİNİ

Abdurrahman Burak GÜHER

DOKTORA TEZİ

Bilgisayar Mühendisliği Anabilim Dalı

Nisan-2021
KONYA
Her Hakkı Saklıdır

TEZ KABUL VE ONAYI

Abdurrahman Burak GÜHER tarafından hazırlanan "***VERİ MADENCİLİĞİ SÜREÇLERİ İLE AKDENİZ BÖLGESİ GÜNEŞ IŞINIM POTANSİYELİNİN TAHMİNİ***" adlı tez çalışması 29/04/2021 tarihinde aşağıdaki jüri tarafından oy birliği / ~~oy çokluğu~~ ile Selçuk Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı'nda DOKTORA TEZİ olarak kabul edilmiştir.

Jüri Üyeleri	**İmza**
Başkan Prof. Dr. Adem Alpaslan ALTUN	……..
Danışman Prof. Dr. Şakir TAŞDEMİR	……..
Üye Doç. Dr. Bülent YANIKTEPE	……..
Üye Dr. Öğr. Üyesi İlker Ali ÖZKAN	……..
Üye Dr. Öğr. Üyesi Onur İNAN	……..

Yukarıdaki sonucu onaylarım.

Prof. Dr. Sait GEZGİN
FBE Müdürü

TEZ BİLDİRİMİ

Bu tezdeki bütün bilgilerin etik davranış ve akademik kurallar çerçevesinde elde edildiğini ve tez yazım kurallarına uygun olarak hazırlanan bu çalışmada bana ait olmayan her türlü ifade ve bilginin kaynağına eksiksiz atıf yapıldığını bildiririm.

DECLARATION PAGE

I hereby declare that all information in this document has been obtained and presented in accordance with academic rules and ethical conduct. I also declare that, as required by these rules and conduct, I have fully cited and referenced all material and results that are not original to this work.

Abdurrahman Burak GÜHER

Tarih: 29.04.2021

ÖZET

DOKTORA TEZİ

VERİ MADENCİLİĞİ SÜREÇLERİ İLE AKDENİZ BÖLGESİ GÜNEŞ IŞINIM POTANSİYELİNİN TAHMİNİ

Abdurrahman Burak GÜHER

Selçuk Üniversitesi Fen Bilimleri Enstitüsü
Bilgisayar Mühendisliği Anabilim Dalı

Danışman: Prof. Dr. Şakir TAŞDEMİR

2021, 166 Sayfa

Jüri
Danışman: Prof. Dr. Şakir TAŞDEMİR
Prof. Dr. Adem Alpaslan ALTUN
Doç. Dr. Bülent YANIKTEPE
Dr. Öğr. Üyesi İlker Ali ÖZKAN
Dr. Öğr. Üyesi Onur İNAN

Bu tez çalışmasında, veri-tahmin ve algoritma-tahmin ilişkisi çerçevesinde seçilerek kullanılan ML algoritmaları ile Akdeniz Bölgesi hedef alınarak Küresel Güneş Işınımı (KGI)'nın tahmin edilmesi amaçlanmıştır. Hedef yerler için yapılan tahmin çalışmalarında Çok Katmanlı İleri Beslemeli Sinir Ağı (ÇKİBSA), Destek Vektör Regresyonu (DVR), K-En Yakın Komşu (K-NN) ve M5 Kuralları ML algoritmaları kullanılmıştır. ML algoritmalarının tahmin performansı büyük oranda kullanılan veri setinin yapısına duyarlı olduğundan, KGI tahmin çalışmalarında nadir kullanılan Veri Madenciliği (VM) süreçleri ile ML algoritmalarının bir arada kullanıldığı tahmin modelleri önerilmiştir.

Çalışmada ilk olarak, VM süreçlerinde kullanılan ön işlem adımları ile her bir ilin ham verileri üzerindeki eksiklikler giderilerek düzenlenmiş ve veri setindeki her değişken için 23,442 adet veri tahmin modellerinde kullanılabilir hale getirilmiştir. Ayrıca VM süreçlerinden KGI üzerinde en etkili olan veri özelliklerinin belirlenmesi sağlanarak, özellik seçim fonksiyonları temelinde girdi veri grupları oluşturulmuştur.

İkinci olarak, tezin ana konusu olan veri-tahmin ilişkisinin değerlendirilmesinde oluşturulan veri grupları kullanılarak, ÇKİBSA ML algoritması ile Akdeniz bölgesinin her bir ili için ayrıntılı tahmin modelleri geliştirilmiş ve tahmin sonuçları birbirleri arasında karşılaştırılmıştır. İller arasında en iyi KGI tahmin sonucu GR3 özellik seçim veri grubu ile Mersin ilinde, R^2 = 0.9668 ve SMAPE = %6.11 bulunmuştur. Akdeniz bölgesi illerinin ÇKİBSA ML algoritması kullanılarak ayrı ayrı KGI potansiyelinin tahmin edilmesinin yanında, 8 ilin verilerinin bir araya getirilerek bölge geneli bir çalışma yapılmış ve iller bazında yapılan modelleme çalışmalarının genel performansının değerlendirilmesi sağlanmıştır. Bölgenin geneli için geliştirilen modelin en iyi tahmin performansı GR3 veri grubunda R^2 = 0.9587 olarak hesaplanırken, 8 ilin R^2 tahmin ortalaması yaklaşık olarak 0.9598 bulunmuştur.

Üçüncü olarak, farklı ML algoritmalarının KGI tahmin performanslarının değerlendirilmesi için aynı enlem üzerinde bulunan iki ilin veri grupları kullanılarak karşılaştırmalı bir çalışma yapılmıştır. Sonuçlar, hedef yerler için, özellik seçim grupları ile geliştirilen tahminlerin kullanılan her bir ML algoritmasında daha başarılı olduğunu ve her iki ilin ışınım potansiyellerinin benzeştiğini ortaya koymuştur. Tahmin modellerinin performansları farklı istatistiksel göstergeler ile karşılaştırıldığında M5 Kuralları ve K-NN için Kahramanmaraş ilinde MAE ölçeğine göre sırasıyla 0.0418 ve 0.0419 değerleri elde edilirken, Isparta ilinde R^2 ölçeğine göre 0.9254 ve 0.9261 sonuçlarına ulaşılmış ve birbirlerine oldukça yakın değerler elde edilmiştir. En düşük tahmin performansı her iki ilde DVR tabanlı LibSVM ile geliştirilen modellerde hesaplanırken, en iyi tahminler ÇKİBSA ML algoritmaları ile elde edilmiştir.

Isparta ve Kahramanmaraş illeri için ÇKİBSA ile geliştirilen modellerden en başarılı sonuçlara MAE ve SMAPE istatistiksel ölçeklerine göre sırasıyla 0.0352-0.0341 ve %7.77-%7.79 değerleri ile ulaşılmıştır.

Çalışma sonucunda, VM süreçlerinin kullanılan ML algoritmalarını KGI tahmininde oldukça iyi bir tahmin performansı seviyesine ulaştırdığı ve ayrıca modelleme süresini kısaltarak modelin yorumlanmasına önemli bir katkı sağladığı görülmüştür. Veri-tahmin ve algoritma-tahmin ilişkisinin değerlendirildiği çalışmaların tamamında en iyi tahmin performansının elde edilmesini sağlayan özellik seçim veri gruplarındaki değişkenlerin dışında kalan meteorolojik özelliklerin tahmin performansına hiçbir etkisinin olmadığı veya performansı düşürdüğü gözlemlenmiştir. Bunula birlikte yapılan modelleme çalışmalarında KGI üzerinde en etkili meteorolojik verinin Güneşlenme Süresi olduğu anlaşılmıştır. Ayrıca Hatay ili hariç yapılan çalışmaların tamamında tüm girdi özelliklerini kullanmanın en iyi tahmin performansına ulaşmayı sağlayamadığı gösterilmiştir.

Anahtar Kelimeler: ÇKİBSA, DVR, Güneş Işınım Tahmini, K-NN, Makine Öğrenme Algoritmaları, M5 Kuralları, Özellik Seçimi, Veri Madenciliği Süreçleri

ABSTRACT

Ph.D THESIS

USING DATA MINING PROCESSES TO THE ESTIMATE OF SOLAR RADIATION POTENTIAL IN THE MEDITERRANEAN REGION

Abdurrahman Burak GÜHER

Selçuk University
Institute of Natural Sciences
Department of Computer Engineering

Advisor: Prof. Dr. Şakir TAŞDEMİR

2021, 166 Pages

Jury

Advisor: Prof. Dr. Şakir TAŞDEMİR
Prof. Dr. Adem Alpaslan ALTUN
Assoc. Prof. Dr. Bülent YANIKTEPE
Asst. Prof. Dr. İlker Ali ÖZKAN
Asst. Prof. Dr. Onur İNAN

In this thesis study, the purpose was to the estimate Global Solar Radiation (KGI) by targeting the Mediterranean Region with ML algorithms that were selected based on data-estimation and algorithm-estimation relations. Multilayer Feedforward Neural Network (ÇKİBSA), Support Vector Regression (DVR), K-Nearest Neighbor (K-NN), and M5 Rules ML Algorithms were used for the estimation studies of target places. Since the predictive performance of ML algorithms is largely sensitive to the structure of the dataset employed in this respect, the estimation models that employ a combination of rarely used Data Mining (VM) processes and ML algorithms were proposed in KGI estimation studies.

In the study, the first step was to eliminate the deficiencies in the raw data of each city with pre-processing steps used in VM processes and to organize these data and 23,442 data were made useable for each variable in the dataset for the estimation models. Also, the input data groups were created based on the basis of feature selection functions after determining the data attributes, which were most effective on KGI from VM processes.

Secondly, detailed prediction models were developed for each city in the Mediterranean Region with the ÇKİBSA ML Algorithm by using the data groups created in the evaluation of the data-estimation relations, which is the main subject of this thesis, and the estimation results were compared with each other. Among the cities, the best KGI estimation result was detected in Mersin with GR3 feature selection data group as R^2 = 0.9668 and SMAPE = 6.11%. In addition to estimation the KGI potential of Mediterranean Region cities separately by using the ÇKİBSA ML Algorithm, a region-wide study was conducted by merging the data of 8 provinces to evaluate the overall performance of the modelling studies on a provincial basis. The best prediction performance of the model that was developed for the Mediterranean Region was calculated as R^2 = 0.9587 in GR3 data group, and the R^2 prediction average of 8 provinces was found to be approximately 0.9598.

Thirdly, a comparative study was conducted by using the data groups of two provinces, which are located on the same latitude, to evaluate the KGI prediction performance of different ML algorithms. The results showed that the predictions that were developed with feature selection groups were more successful for target locations in each ML algorithm employed, and the solar radiation potentials of both provinces were similar. When the performance of the prediction models was compared with different statistical indicators, for M5 Rules and K-NN, 0.0418 and 0.0419 values were obtained respectively according to the MAE scale in Kahramanmaraş; and 0.9254 and 0.9261 results were obtained according

to the R^2 scale in Isparta, and these values were very close to each other. The lowest prediction performance was calculated in models that were developed with the SVR-based LibSVM in both provinces, and the best predictions were obtained with the ÇKİBSA ML algorithms. The most successful results among the models that were developed with ÇKİBSA for Isparta and Kahramanmaraş were obtained with 0.0352-0.0341 and 7.77%-7.79%, respectively according to the MAE and SMAPE statistical scales.

As a result, it was concluded that the VM processes bring the ML Algorithms to a fairly good level in terms of estimation performance in KGI, and also contribute to the interpretation of the model at significant levels by shortening the modelling time. In all previous studies in which the data-estimation and algorithm-estimation relations were evaluated, it was found that the meteorological attributes outside the variables in the feature selection data groups, which ensured the best estimate performance, had no effects on -or decreased the performance of- the estimation performance. However, it was also understood in the modelling studies that the most effective meteorological data on KGI was Sunshine Duration. Also, it was concluded that using all the input features in all studies -except for Hatay- does not achieve the best estimate performance.

Keywords: ÇKİBSA, Data Mining Processes, DVR, Feature Selection, K-NN, Machine Learning Algorithms, M5 Rules, Solar Radiation Estimation

ÖNSÖZ

Son yıllarda dünyada ve ülkemizde yenilenebilir enerji kaynakları arasında en fazla yatırımın yapıldığı alan olan güneş enerjisi, temiz, sürdürülebilir ve çevreci olmasıyla dikkatleri üzerine çekmektedir. Herhangi bir yere güneş enerji santralli kurulmadan önce yatırımın planlanması için güneş ışınım potansiyellerinin hesaplanması gerekmektedir. Meteorolojik ölçümlerde yaşanan birtakım problemler sebebiyle tahminsel yöntemler geliştirilmiştir. Güneş ışınımının gerçek ölçüm değerlerine oldukça yakın tahmin edilmesi en önemli problemlerin başında gelmektedir. Bu tez çalışmasında tahmin yöntemlerinde kullanılan veri miktarları ve değişken sayılarının artmasıyla, güneş ışınımının yüksek doğrulukta, hızlı ve yorumlanabilir bir şekilde tahmin edilmesini sağlamak amacıyla VM süreçlerine dayanan ML tabanlı tahmin modelleri geliştirilip, uygulama sonuçları farklı yöntem ve tekniklerle değerlendirilmiştir. Önerilen yöntem sayesinde eldeki veri setinin yapısına uygun tahmin modellerinin geliştirilmesi sağlanarak, uygulama eksikliklerin önüne geçilmesi hedeflenmiş, daha kesin değerlendirmelerin yapılaması sağlanmıştır.

Yapılan çalışma 3 bölümde incelenmiştir. İlk bölümde VM yöntemleri kullanılarak bölge genelinde toplanan saatlik verilere ön işlem adımları uygulanarak veri grupları oluşturulmuştur. İkinci olarak veri-tahmin ilişkisi temelinde il bazında ve bölge genelinde tahmin modelleri geliştirilmiş ve değerlendirilmiştir. Üçüncü olarak, farklı ML algoritmaların tahmin performansları algoritma-tahmin ilişkisi çerçevesinde karşılaştırılmıştır.

Yapılan çalışmada, meteorolojik ölçüm istasyonlarından kaynaklanan ölçüm problemleri nedeniyle, sağlıklı ve güncel güneş ışınım ölçüm verilerinin bölge genelinde talep edilip ve toplanmasında çok büyük zorluklarla karşılaşılmıştır. Ayrıca elde edilen verilerin düzenlenmesi ve amaca uygun seçim süreçlerinin hedeflenenden oldukça fazla bir süreç alması da karşılaşılan bir başka önemli zorluktur. Veri temininde karşılaştığım sorunları aşmamda teknik desteğinden faydalandığım MGM müdürlüğü teknik personeli Ali SARGIN Beye teşekkürü bir borç bilirim.

Doktora eğitimim sürecince bilimsel özeni ile bana yol gösteren ve tüm süreçlerde desteklerini esirgemeyen danışmanlığımı yapan Teknoloji Fakültesi Dekanı Sayın hocam Prof. Dr. Şakir TAŞDEMİR'e sonsuz teşekkür ve minnetimi sunarım. Doktora eğitime başlamamdan tezimin tamamlanmasına kadar geçen sürede üzerimde büyük emekleri olan, desteğini her zaman yanımda hissettiğim ve tez jürimde bulunarak benden değerli görüş ve katkılarını esirgemeyen kıymetli hocam Doç. Dr. Bülent YANIKTEPE'ye müteşekkirim. Ayrıca tez jürimde bulunan Prof. Dr. Adem Alpaslan ALTUN hocama doktora çalışmamda yaptığı öneriler ve eleştirilerinden dolayı teşekkürü bir borç bilirim. Son olarak bu çalışmanın hazırlanması süreci boyunca maddi manevi desteğini esirgemeyen aileme, değerli eşim Doç. Dr. Ebru GÜHER'e, sevgileriyle her zaman bana güç veren kızım Elif DURU ve oğlum Ali TUNA'ya sonsuz teşekkürlerimi sunarım.

Abdurrahman Burak GÜHER
KONYA-2021

İÇİNDEKİLER

ÖZET ..... **iv**

ABSTRACT ..... **vi**

ÖNSÖZ ..... **viii**

İÇİNDEKİLER ..... **ix**

ŞEKİLLER LİSTESİ ..... **xi**

ÇİZELGELER LİSTESİ ..... **xiv**

SİMGELER VE KISALTMALAR ..... **xv**

1. GİRİŞ **1**

1.1. Güneş Işınımını Ölçme Yöntemleri 4
1.2. Güneş Işınımı Tahmin Yöntemleri 5
1.2.1. Deneysel yöntemler 5
1.2.2. İstatistiksel yöntemler 7
1.2.3. Yapay zekâ yöntemleri 9
1.2.3.1. Makine öğrenme yöntemleri 10
1.3. Veri Madenciliği 12
1.3.1. Amaca uygun özelliklerin seçimi 16
1.4. Çalışmanın Amacı ve Önemi 18

2. KAYNAK ARAŞTIRMASI **23**

2.1. ML Yöntemleri Temelli Çalışmalar 23
2.1.1. Algoritma tabanlı modelleme 23
2.1.2. Veri tabanlı modelleme 30
2.2. VM Yöntemleri Temelli Çalışmalar 38

3. MATERYAL VE YÖNTEM **43**

3.1. Türkiye'de Güneş Işınım Potansiyeli 43
3.2. Çalışma Alanı ve Verilerin Seçimi 45
3.3. ML Tabanlı Tahmin Algoritmaları 55
3.3.1. YSA algoritması 58
3.3.2. DVR algoritması 63
3.3.2.1 LibSVM kütüphanesi 65
3.3.3. K-NN algoritması 66
3.3.4. M5 kuralları algoritması 67
3.4. İstatistiksel Performans Değerlendirme Metrikleri 69

4. VERİ MADENCİLİĞİ SÜREÇLERİ **72**

4.1. Tahmin Modellerinde Kullanılacak Verilerin Ön İşlem Süreçleri 72

4.2. Girdi Verilerinden Optimal Özelliklerin Seçim Süreçleri 78

5. GELİŞTİRİLEN MODELLERİN BULGULARI VE TARTIŞMA 85

5.1. ÇKİBSA Algoritması ile Tahmin Modellerinin Geliştirilmesi..... 85
5.1.1. Adana ilinin tahmin bulguları 88
5.1.2. Antalya ilinin tahmin bulguları..... 92
5.1.3. Burdur ilinin tahmin bulguları 96
5.1.4. Hatay ilinin tahmin bulguları 99
5.1.5. Isparta ilinin tahmin bulguları..... 103
5.1.6. Kahramanmaraş ilinin tahmin bulguları 107
5.1.7. Mersin ilinin tahmin bulguları 110
5.1.8. Osmaniye ilinin tahmin bulguları 113
5.1.9. ÇKİBSA tahmin modellerinin değerlendirilmesi 117
5.2. Farklı ML Algoritmaları ile Geliştirilen Modellerin Karşılaştırmalı Analizi.... 122
5.2.1. DVR algoritması ile geliştirilen modellerin tahmin bulguları 123
5.2.2. K-NN algoritması ile geliştirilen modellerin tahmin bulguları 126
5.2.3. M5 kuralları algoritması ile geliştirilen modellerin tahmin bulguları 127
5.2.4. Hedef yerlerin tahmin bulgularının değerlendirilmesi..... 128

6. SONUÇLAR VE ÖNERİLER 133

6.1 Sonuçlar 133
6.2 Öneriler 137

KAYNAKLAR 139

ÖZGEÇMİŞ 148

ŞEKİLLER LİSTESİ

Şekil 1.1. 2010-2019 yılları arası PV tabanlı santrallerinin kapasite miktarları 3
Şekil 1.2. Piyranometre cihazının temel yapısı 5
Şekil 1.3. Veri Madenciliğinin bilgisayar bilimi alanları ile ilişkisel yapısı 14
Şekil 1.4. Veri madenciliği süreçleri işlem prosesleri 15
Şekil 1.5. Özellik seçim sürecinin çalışma prosedürü 17
Şekil 1.6. Filtre, Sarmalayıcı ve Gömülü özellik seçim yöntemlerinin işlem adımları .. 18
Şekil 1.7 Yapılacak çalışmanın metodolojik çerçevesi 22

Şekil 3.1. Türkiye'nin güneş enerjisi potansiyeli atlası 44
Şekil 3.2. Akdeniz bölgesinde seçilen iller ve meteorolojik ölçüm istasyonlarının enlem koordinatları 46
Şekil 3.3. Adana ilinin 2002-2006 yılları arasında ölçülen KGI'nın 3-Boyutlu yıllık dağılım grafikleri 48
Şekil 3.4. Antalya ilinin 2002-2006 yılları arasında ölçülen KGI'nın 3-Boyutlu yıllık dağılım grafikleri 49
Şekil 3.5. Burdur ilinin 2002-2006 yılları arasında ölçülen KGI'nın 3-Boyutlu yıllık dağılım grafikleri 50
Şekil 3.6. Hatay ilinin 2002-2006 yılları arasında ölçülen KGI'nın 3-Boyutlu yıllık dağılım grafikleri 51
Şekil 3.7. Isparta ilinin 2002-2006 yılları arasında ölçülen KGI'nın 3-Boyutlu yıllık dağılım grafikleri 52
Şekil 3.8. Kahramanmaraş ilinin 2002-2006 yılları arasında ölçülen KGI'nın 3-Boyutlu yıllık dağılım grafikleri 53
Şekil 3.9. Mersin ilinin 2002-2006 yılları arasında ölçülen KGI'nın 3-Boyutlu yıllık dağılım grafikleri 54
Şekil 3.10. Osmaniye ilinin 2002-2006 yılları arasında ölçülen KGI'nın 3-Boyutlu yıllık dağılım grafikleri 55
Şekil 3.11. Yaygın kullanılan ML algoritmaları 57
Şekil 3.12. Basit tek katmalı algılayıcının yapısı 58
Şekil 3.13. Tahmin modellerinde geliştirilen ÇKİBSA mimari yapısı 61
Şekil 3.14. DVR tabanlı geliştirilen modelin mimari yapısı 64

Şekil 4.1. Filtre yaklaşımının algoritmik yapısının şematik gösterimi 80

Şekil 5.1. ÇKİBSA algoritmasına göre geliştirilen modellerin akış şeması 86
Şekil 5.2. Adana ili için geliştirilen en iyi modelin ÇKİBSA performans grafiği 90
Şekil 5.3. Adana ili veri gruplarına göre ölçülen ve tahmin edilen değerler arasındaki hata kutu grafiği 91
Şekil 5.4. Adana ili test veri gruplarına göre geliştirilen ÇKİBSA modellerinin KGI tahmin dağılım grafikleri 91
Şekil 5.5. Adana ili 2003 yılı Mayıs ayı dört günlük ölçülen ve tahmin edilen saatlik verilerin karşılaştırmalı performans grafiği 92
Şekil 5.6. Antalya ili için geliştirilen en iyi modelin ÇKİBSA performans grafiği 94
Şekil 5.7. Antalya ili veri gruplarına göre ölçülen ve tahmin edilen değerler arasındaki hata kutu grafiği 94
Şekil 5.8. Antalya ili test veri gruplarına göre geliştirilen ÇKİBSA modellerinin KGI tahmin dağılım grafikleri 95
Şekil 5.9. Antalya ili 2006 yılı Temmuz ayı dört günlük ölçülen ve tahmin edilen saatlik verilerin karşılaştırmalı performans grafiği 96

Şekil 5.10. Burdur ili için geliştirilen en iyi modelin ÇKİBSA performans grafiği 97
Şekil 5.11. Burdur ili veri gruplarına göre ölçülen ve tahmin edilen değerler arasındaki hata kutu grafiği 98
Şekil 5.12. Burdur ili test veri gruplarına göre geliştirilen ÇKİBSA modellerinin KGI tahmin dağılım grafikleri 98
Şekil 5.13. Burdur ili 2002 yılı Nisan ayı dört günlük ölçülen ve tahmin edilen saatlik verilerin karşılaştırmalı performans grafiği 99
Şekil 5.14. Hatay ili için geliştirilen en iyi modelin ÇKİBSA performans grafiği..... 101
Şekil 5.15. Hatay ili veri gruplarına göre ölçülen ve tahmin edilen değerler arasındaki hata kutu grafiği 101
Şekil 5.16. Hatay ili test veri gruplarına göre geliştirilen ÇKİBSA modellerinin KGI tahmin dağılım grafikleri 102
Şekil 5.17. Hatay ili 2002 yılı Nisan ayı dört günlük ölçülen ve tahmin edilen saatlik verilerin karşılaştırmalı performans grafiği 103
Şekil 5.18. Isparta ili için geliştirilen en iyi modelin ÇKİBSA performans grafiği 105
Şekil 5.19. Isparta ili test veri gruplarına göre geliştirilen ÇKİBSA modellerinin KGI tahmin dağılım grafikleri 105
Şekil 5.20. Isparta ili veri gruplarına göre ölçülen ve tahmin edilen değerler arasındaki hata kutu grafiği 106
Şekil 5.21. Isparta ili 2002 yılı Mayıs ayı dört günlük ölçülen ve tahmin edilen saatlik verilerin karşılaştırmalı performans grafiği 106
Şekil 5.22. Kahramanmaraş ili için geliştirilen en iyi modelin ÇKİBSA performans grafiği 108
Şekil 5.23. Kahramanmaraş ili veri gruplarına göre ölçülen ve tahmin edilen değerler arasındaki hata kutu grafiği 108
Şekil 5.24. Kahramanmaraş ili test veri gruplarına göre geliştirilen ÇKİBSA modellerinin KGI tahmin dağılım grafikleri 109
Şekil 5.25. Kahramanmaraş ili 2003 yılı Haziran ayı dört günlük ölçülen ve tahmin edilen saatlik verilerin karşılaştırmalı performans grafiği 110
Şekil 5.26. Mersin ili için geliştirilen en iyi modelin ÇKİBSA performans grafiği 111
Şekil 5.27. Mersin ili veri gruplarına göre ölçülen ve tahmin edilen değerler arasındaki hata kutu grafiği 112
Şekil 5.28. Mersin ili test veri gruplarına göre geliştirilen ÇKİBSA modellerinin KGI tahmin dağılım grafikleri 112
Şekil 5.29. Mersin ili 2005 yılı Nisan ayı dört günlük ölçülen ve tahmin edilen saatlik verilerin karşılaştırmalı performans grafiği 113
Şekil 5.30. Osmaniye ili için geliştirilen en iyi modelin ÇKİBSA performans grafiği 115
Şekil 5.31. Osmaniye ili test veri gruplarına göre geliştirilen ÇKİBSA modellerinin KGI tahmin dağılım grafikleri 115
Şekil 5.32. Osmaniye ili veri gruplarına göre ölçülen ve tahmin edilen değerler arasındaki hata kutu grafiği 116
Şekil 5.33. Osmaniye ili 2004 yılı Temmuz ayı dört günlük ölçülen ve tahmin edilen saatlik verilerin karşılaştırmalı performans grafiği 116
Şekil 5.34. Akdeniz bölgesi GR3 ve GR7 veri gruplarına göre geliştirilen ÇKİBSA modellerinin KGI tahmin dağılım grafikleri 121
Şekil 5.35. Akdeniz bölgesi GR3 ve GR7 veri gruplarına göre ölçülen ve tahmin edilen değerler arasındaki hata kutu grafiği 121
Şekil 5.36. Farklı ML algoritmaları kullanılarak KGI tahmin süreçlerinin şematik gösterimi 123

Şekil 5.37. Klasik DVR ile geliştirilen modellerin (a) Kahramanmaraş (b) Isparta ileri girdi veri gruplarına göre tahmin performansları. 124
Şekil 5.38. Klasik DVR ve LibSVM'ye göre geliştirilen en başarılı modellerin (a) Isparta, (b) Kahramanmaraş illerine göre karşılaştırmalı dağılım grafikleri 125
Şekil 5.39. K-NN tabanlı en başarılı modellerin (a) Kahramanmaraş GR2, (b) Isparta GR3 girdi verilerine göre dağılım grafikleri 127
Şekil 5.40. M5 Kuralları tabanlı en başarılı modellerin (a) Isparta GR3, (b) Kahramanmaraş GR5 girdi gruplarına göre dağılım grafikleri 128
Şekil 5.41. Veri gruplarına göre ML algoritmaları kullanılarak geliştirilen en iyi modellerin (a) Kahramanmaraş, (b) Isparta illeri için MAE performans karşılaştırmaları 130
Şekil 5.42. Veri gruplarına göre ML algoritmaları kullanılarak geliştirilen en iyi modellerin (a) Kahramanmaraş, (b) Isparta illeri için SMAPE performans karşılaştırmaları 130
Şekil 5.43. Isparta ili 2004 yılı Temmuz ayı beş günlük girdi verileri kullanılarak ölçülen ve tahmin edilen 4 farklı ML modellerinin karşılaştırmalı performans grafiği 131
Şekil 5.44. Kahramanmaraş ili 2004 yılı Temmuz ayı beş günlük girdi verileri kullanılarak ölçülen ve tahmin edilen 4 farklı ML modellerinin karşılaştırmalı performans grafiği 131

ÇİZELGELER LİSTESİ

Çizelge 3.1. Türkiye'nin bölgelere göre güneş ışınım potansiyelleri 45
Çizelge 3.2. Çalışmada seçilen her bir ilin coğrafik ve meteorolojik verilerinin spesifik özellikleri 47

Çizelge 4.1. Akdeniz Bölgesi illerinin ham meteorolojik verilerinin karakteristik özellikleri 74
Çizelge 4.2. Eğitim ve test verilerinin seçimine göre tahmin performansları 77
Çizelge 4.3. Veri ön işleme süreci sonunda her bir il için düzenlenen veri setinin nihai özellikleri 78
Çizelge 4.4. Mersin ili veri setine uygulanan 29 farklı uygulamanın seçim sonuçları ... 82
Çizelge 4.5. Filtre yaklaşımı temelli özellik seçim süreci sonunda geliştirilen uygulamanın ekran çıktısı 82
Çizelge 4.6. Akdeniz Bölgesi illerinin özellik seçim fonksiyonlarına göre seçilen veri özellikleri 83
Çizelge 4.7. Akdeniz Bölgesi illerinin tahmin modellerinde kullanılmak üzere oluşturulan nihai girdi veri grupları ve özellik sayıları 84

Çizelge 5.1. Adana ili veri gruplarına göre geliştirilen en başarılı modellerin eğitim ve test tahmin sonuçları 89
Çizelge 5.2. Antalya ili veri gruplarına göre geliştirilen en başarılı modellerin eğitim ve test tahmin sonuçları 93
Çizelge 5.3. Burdur ili veri gruplarına göre geliştirilen en başarılı modellerin eğitim ve test tahmin sonuçları 97
Çizelge 5.4. Hatay ili veri gruplarına göre geliştirilen en başarılı modellerin eğitim ve test tahmin sonuçları 100
Çizelge 5.5. Isparta ili veri gruplarına göre geliştirilen en başarılı modellerin eğitim ve test tahmin sonuçları 104
Çizelge 5.5.6. Kahramanmaraş ili veri gruplarına göre geliştirilen en başarılı modellerin eğitim ve test tahmin sonuçları 107
Çizelge 5.7. Mersin ili veri gruplarına göre geliştirilen en başarılı modellerin eğitim ve test tahmin sonuçları 111
Çizelge 5.8. Osmaniye ili veri gruplarına göre geliştirilen en başarılı modellerin eğitim ve test tahmin sonuçları 114
Çizelge 5.5.9. Akdeniz bölgesinin en başarılı KGI modellerinin tahmin performansı sonuçları 117
Çizelge 5.10. Akdeniz bölgesi için belirlenen özellik seçim grupları 120
Çizelge 5.11. Akdeniz bölgesi veri gruplarına göre geliştirilen en başarılı modellerin eğitim ve test tahmin sonuçları 120
Çizelge 5.12. Her iki ilin girdi veri gruplarına göre LibSVM'ye göre geliştirilen en başarılı modellerin tahmin sonuçları 125
Çizelge 5.13. Hedef yerlerin veri gruplarına göre K-NN algoritması kullanılarak geliştirilen en başarılı iki modelinin tahmin sonuçları 126
Çizelge 5.14. M5 Kuralları algoritması kullanılarak geliştirilen en başarılı modellerin tahmin sonuçları 127
Çizelge 5.15. Her iki ilin farklı ML algoritmaları kullanılarak geliştirilen en başarılı KGI tahmin modellerinin karşılaştırması 129
Çizelge 5.16. Literatürde saatlik ortalama KGI tahmin çalışmaları ile mevcut çalışmanın karşılaştırılması 132

SİMGELER VE KISALTMALAR

Simgeler

α_a ve α_a^*	: Lagrange Çarpanları
e^{x_i}	: Doğal Logaritma
ε	: Epsilon
$\overline{X}$	: Girdi Verilerinin Ortalaması
$\varphi(x)$	: Haritalama Fonksiyonu
μ_i	: Ortalama
$\overline{O}$	: Ölçüm Ortalaması
$\overline{P}$	: Tahmin Ortalaması
°	: Derece
°C	: Derece Santigrat
b_i	: Eşik Değer-Bias
C	: Sabit
Cal	: Kalori
Cm^2	: Santimetre Kare
e	: Ağ Hatası
G	: Ağ Gradyanı
GW	: Gigawatt
GWh	: Gigawat Saat
H	: Hessian Matrisi
hPa	: Hectopascal
I	: Birim Matris
J	: Ağırlık
K	: Komşuluk Katsayısı
$K(x_a,x)$	: Çekirdek-Kernel
km^2	: Kilometre Kare
kWh	: Kilowatt Saat
m^2	: Metre Kare
MW	: Megawatt
N_i	: Hesaplanan Net Girdi
O_i	: Ölçülen Değer
P_i	: Tahmin Edilen Değer
S - σ_i	: Standart Sapma
Σ	: Toplam Fonksiyonu
W	: Watt
W_{ij}	: Ağırlık
x_a	: Girdi Vektörü
X_j	: Girdi
y_a	: Çıktı Vektörü
Y_i	: Çıktı
μ	: Kombinasyon Katsayısı
φ	: Aktivasyon Fonksiyonu

Kısaltmalar

ANFIS : Uyarlanabilir Nöro-Bulanık Çıkarım Sistemi
AR : Otoregresif
ARIMA : Otoregresif Entegre (I) Hareketli Ortalama
ARMA : Otoregresif Hareketli Ortalama
B : Basınç
CART : Regresyon Ağaçları
CFNN : Basamak Korelasyon Sinir Ağı
CfsSE : Korelasyona dayalı Özellik Seçimi Alt Kümesi Değerlendiricisi-Corelation-based Future Selection Subset Evaluator
CGP : Pola-Ribiere Konjugat Gradyan
ClassAE : Sınıflandırıcı Öznitelik Değerlendiricisi-Classifier Attribute Evaluator
CorrAE : Korelasyon Özellik Değerlendiricisi-Corelation Attribute Evaluator
CSE : Sınıflandırıcı Alt Küme Değerlendiricisi-Classifier Subset Evaluator
ÇKİBSA : Çok Katmanlı İleri Beslemeli Sinir Ağı
DL : Derin Öğrenme
DNN : Dalgacık Dönüşümü Sinir Ağı
DVM : Destek Vektör Makinası
DVR : Destek Vektör Regresyonu
ELMNN : Elman Geri Yayılım Sinir Ağı
EML : Aşırı Öğrenme Makinesi
GA : Genetik Algoritma
GEPA : Türkiye Güneş Enerjisi Potansiyeli Atlası
GR : Veri Grubu
GRNN : Genelleştirilmiş Regresyon Sinir Ağı
KGI : Küresel Güneş Işınımı
K-NN : K-En Yakın Komşu
LM : Levenberg-Marquardt Algoritması
MA : Hareketli Ortalama
MAE : Ortalama Mutlak Hata
MAPE : Ortalama Mutlak Yüzde Hata
MBE : Ortalama Sapma Hatası
MGM : Meteoroloji Genel Müdürlüğü
ML : Makine Öğrenme
MLFFBPN : Çok Katmanlı İleri Beslemeli Geri Yayılım Ağı
MLP : Çok Katmanlı Algılayıcılar
MNLR : Çoklu Doğrusal Olmayan Regresyon
N : Nem
NARX : Doğrusal Olmayan Dışsal Girdili Otoregresif Ağ
NOR-P : Normalize Polinomial
POL : Polinomial
PV : Fotovoltaik-Güneş Paneli
R : Korelasyon Katsayısı
R^2 : Belirtme Katsayısı
RAE : Relief Özellik Değerlendiricisi-Relief Attribute Evaluator
RBF : Gaussian Radial Basis Functions
RBFNN : Radyal Temelli Fonksiyon Sinir Ağı
RF : Rastgele Orman
RH : Rüzgâr Hızı

RMSE	: Kök Ortalama Kare Hatası
RNN	: Tekrarlayan Sinir Ağı
S	: Sıcaklık
SCG	: Ölçekli Eşlenik Gradyan
SMAPE	: Simetrik Ortalama Mutlak Yüzde Hata
VM	: Veri Madenciliği
WEKA	: Waikato Environment for Knowledge Analysis
WSE	: Sarmalayıcı Alt Küme Değerlendiricisi-Wrapper Subset Evaluator
YSA	: Yapay Sinir Ağları
YZ	: Yapay Zekâ

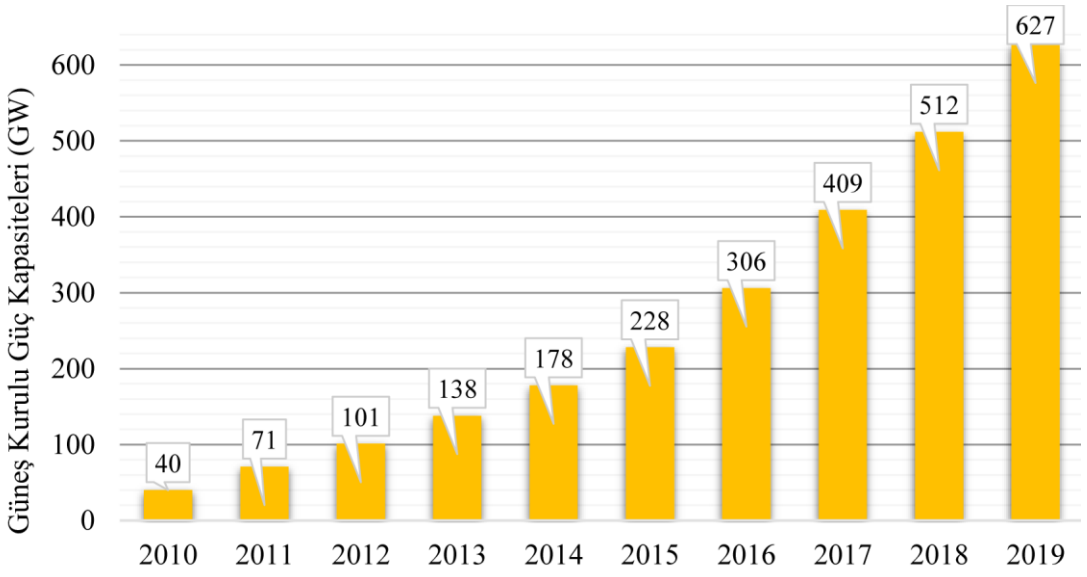
1. GİRİŞ

Enerji, ülkeler için en önemli iktisadi girdi kalemi ve insanların sosyo-ekonomik hayatın sürdürülebilmesi için önemli yapı taşıdır. Ülkelerin hali hazırda var olan enerji kaynaklarına olan bağımlılığı ve yeni kaynak arayışları büyük oranda artmış ve hatta farklı ülkeler ile uluslararası bir sorun haline gelmiş durumdadır. Kalkınmada da etkin bir parametre olan enerji, sanayi, teknolojik gelişmeler ve artan nüfus ile birlikte hızla artmaktadır. Günümüzde gelişmişliğin en temel göstergelerinden olan artan kişi başı enerji tüketim miktarları, güvenli, yeterli miktarda, ucuz ve temiz enerji üretmek için ülkeleri büyük oranda zorlamaktadır. Özellikle her ülkenin enerjiye olan gereksinimini karşılamada yeterli düzeyde enerji kaynaklarına sahip olmaması, enerji arzında ülkeleri alternatif kaynaklara yönelmeye zorlamaktadır. Bu sebeple ülkeler, enerji ihtiyacını karşılamada konvansiyonel enerji kaynakları olarak adlandırılan karbon temelli kaynakların yerine, güneş, rüzgâr, hidrolik, biyo, hidrojen, jeotermal, gel-git gibi yenilenebilir enerji kaynaklarını tercih etmektedir. Bu yönelimde, karbon temelli enerji kaynaklarının tükenebilir oluşu, çevreye olan olumsuz etkileri, dışa bağımlılığın gittikçe artması ve devletler arasındaki çözülemez jeopolitik sorunlar oldukça etkili olmaktadır. Bu sebeple, hâlihazırda kullanılan karbon temelli birincil enerji kaynaklarına ilaveten, yenilenebilir enerji kaynaklarına yapılacak yatırımların gerekliliği büyük önem taşımaktadır (Bayrak ve Esen, 2014).

Dünya genelinde yenilenebilir enerji kaynaklarına yönelim son yıllarda artan bir ivme kazanmıştır. Bu ivmelenmede küresel ısınma, kaynak bolluğu, yatırım maliyetlerinin azalması, yenilenebilir teknolojilerdeki yenilenmeler ve sosyo-ekonomik hayata yaptığı olumlu katkılar etkili olmaktadır. Ülkelerin doğru yatırım ve teşvik politikalarını desteklemesi ile yenilenebilir enerjinin yakın bir gelecekte toplam küresel enerji talebinin üçte ikisini karşılayabilir bir duruma gelmesi öngörülmektedir (Gielen ve ark., 2019). Küresel ölçekte gelişmiş veya gelişmekte olan ülkelerin tamamında yenilenebilir enerjiden elde edilen elektrik enerjisi miktarlarındaki artış dikkat çekmektedir. Özellikle şebekeye bağlı (On-Grid) olarak kurulan, yenilenebilir enerji santrallerinin maksimum güç kapasitesi 2010 yılında yaklaşık 1,224 GW iken, 2019 yılında yaklaşık 2,533 GW'lara yükseldiği görülmektedir. Geçen 10 yıllık periyotta yaklaşık %107'lik bir artış sağlanmıştır. Bir önceki yıl olan 2018 yılı ile karşılaştırıldığında yaklaşık %8'lik bir büyüme oranı yakalanmıştır (IRENA, 2020). Ülkelerin enerji ihtiyaçları arttıkça çevre ile dost enerji kaynaklarına yönelimin her

geçen sene, bir önceki seneye kıyasla hızlı bir biçimde büyüdüğü gözlemlenmektedir. Özellikle gelişmekte olan ülkelerde bu artışın daha dikkat çekici bir boyuta ulaştığı bilinmektedir.

Yenilenebilir enerji kaynakları içerisinde bulunan güneş enerjisi son yıllarda ülkelerin en fazla yatırım yaptığı kaynakların başında gelmektedir. Elektrik üretiminde gün geçtikçe kritik bir rol oynayan güneş enerjisi, fosil yakıtlarına kıyasla, temiz, sınırsız ve sürdürülebilir olması nedeniyle ülkelerin dikkatini çeken en umut verici yenilenebilir enerji kaynaklarından birisi olmuştur. Güneş enerjisinden elektrik elde etmek amacıyla yaygın olarak kullanılan sistemleri Fotovoltaik (PV-Güneş Paneli) ve Termal (T) tabanlı sistemler olarak sınıflandırmak mümkündür. Her ikisinin hibrit olarak çalıştığı PV-T sistemleri de kullanılmaktadır. PV sistemlerde enerji doğrudan güneşten sağlanırken, Termal sistemlerde dolaylı yoldan elektrik enerjisine dönüştürülmektedir (Singh, 2013). Elektrik üretiminde PV-tabanlı sistemler dünyada en yoğun kullanılan yöntemlerin başında gelmektedir. 2019 sonu itibari ile dünyada elektrik üretiminin %72.7'si karbon temelli yakıtlardan %27.3 ise yenilenebilir enerji kaynaklarından elde edildiği tahmin edilmektedir. Güneş panelleri kullanılarak elde edilen enerjinin küresel üretime katkısı %2.8 civarındadır. Dünyada, son 10 yıllık periyotta PV tabanlı güneş kurulu güç kapasiteleri Şekil 1.1'de verilmiştir. Sadece 2019 yılında 115 GW'lık yeni güneş enerji santrali kurularak toplamda 627 GW kurulu güç kapasitesi ile güneş enerjisi diğer yenilenebilir enerji kaynakları arasında en fazla yatırım yapılan alan olarak lider konumda bulunmaktadır. Ülkeler bazında bakıldığında Çin, Japonya ve Amerika Birleşik Devletleri en fazla kurulu güce sahip ülkeler olarak görülmektedir. Bu ülkeler arasında Çin, bir önceki yıla kıyasla %26'lık bir artışla 2019 yılında yaklaşık 205 GW'lık kurulu güç kapasitesi ile güneş enerji pazarında önemli bir aktör haline gelmiş durumdadır (REN21, 2020). Güneş enerjisinin dünya genelinde en fazla dikkat çeken yenilenebilir kaynak olmasında, güneş enerjisi alanında gerçekleştirilen teknolojik gelişmelerin artması ve kurulum maliyetlerinin oldukça azalması büyük oranda etkili olmuştur.



Şekil 1.1. 2010-2019 yılları arası PV tabanlı santrallerinin kapasite miktarları

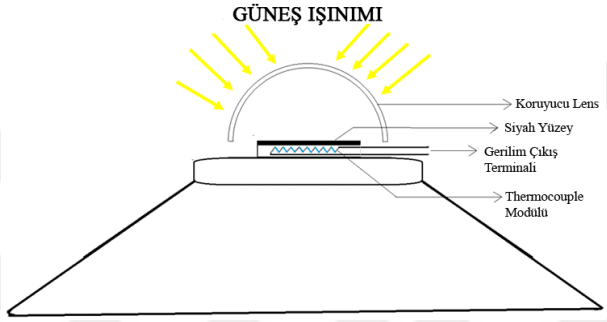
Coğrafi konum olarak Türkiye, dünyadaki birçok ülkeye kıyasla güneş enerjisi potansiyeli bakımından oldukça verimli bir ülkedir. Son yıllarda, güneş enerjisinden faydalanmak için önemli yatırımlar yapılmakta ve PV maliyetlerini düşürmek amacıyla teknoloji transferi yapılarak, üretim üssü olmak hedeflenmektedir. Türkiye'de güneş enerjisi kullanılarak üretilen elektrik enerjisi miktarı 2014 yılı ve öncesinde yok denebilecek kadar azken, 2018 yılından sonra hızlıca armış ve üretimdeki payı %2.6 seviyelerine, 2019'da ise yaklaşık olarak %3.5 seviyesine yükselmiştir. 2018 yılı içinde lisanlı olarak hizmet veren büyük ölçekli güneş enerji santrali sayısı 9 iken, bu sayı 2019 yılında 17 olmuştur. Lisansız olarak (5 MW ve altı) 2018 yılında kurulan santral sayısı toplam 5,859 iken, 2019 yılında %17.5 artış göstererek 6,884'e yükselmiştir. Toplam kurulu güneş gücü 2018 yılında yaklaşık 5,063 MW iken, 2019 yılında yaklaşık 5,995 MW seviyesine ulaşmıştır. Bu santrallerden üretilen elektrik enerjisi 2018 yılında yaklaşık 7,800 GWh iken 2019 yılında yaklaşık 10,000 GWh seviyelerine yükselerek ülkenin toplam elektrik üretiminde önemli bir kaynak olarak hızla ivmelenmektedir. Fosil kaynağı olarak çok zengin olmayan Türkiye'de, elektrik üretiminin yaklaşık %51'i bu tip kaynaklardan sağlandığı düşünüldüğünde, güneş enerjisinin cari açığı ve dışa bağımlılığı azaltacak, sürdürülebilir ve çevreci politikalarla desteklenmesi gerekmektedir. Bu sebeple, Türkiye'nin 2023 enerji hedeflerinde güneş enerjisi kurulu gücünü 10 GW seviyelerine çıkartmak ve devlet destekli büyük ölçekli yatırımlar hedeflenmektedir (Turkey_Investment_Office, 2019; TMMOB, 2020). Güneş potansiyeli olarak oldukça yüksek olan ülkemizin, bu alanda sürdürdüğü yatırım ve

politikalar oldukça yeni olmasına rağmen, dünyada gelişmiş ve gelişmekte olan ülkelerle beraber bu yarıştan kopması kabul edilemez bir gerçeklik olacaktır. Çünkü ülkeler geleceğe dönük güneş enerjisi hedeflerini önemli oranlarda arttırmışlardır. Bunun en önemli sebebi yenilenebilir enerji kaynaklarına olan yoğun ilgi ve güneşin diğer kaynaklara nazaran son yıllardaki popülerliğidir. Ayrıca, bilim çevreleri, güneş enerjisinden çok iyi yararlanmak ve teknolojisindeki gelişmeleri arttırabilmek amacıyla literatürdeki gelişmelere ilaveten yeni tekniklere yönelmeye mecburdurlar.

1.1. Güneş Işınımını Ölçme Yöntemleri

Seçilen coğrafik bir alanda, güneş tabanlı bir sistemin tasarımı ve kurulumundan önce en önemli girdi değeri olarak güneş ışınım verisinin ölçülmesi gerekmektedir. Güneş tabanlı santrallerin kurulumundan önce bu değerin ölçümü ve hesaplanması yatırımın şeklini belirlemedeki en önemli parametredir. Dünya atmosferinin en üst katmanında, dik bir uzay düzlemi üzerine doğrudan düşen güneş enerjisi akı yoğunluğunun miktarı güneş sabiti olarak adlandırılır ve ölçüm değeri yaklaşık olarak 1,367 W/m^2 olarak kabul edilmektedir. Güneş sabiti hatasız bir biçimde dünya atmosferinin dışındaki uydular tarafından ölçülebilmektedir. Bu değer atmosferin katmanlarından geçerken emilme, dağılma veya yansıtılma gibi birtakım olaylara maruz kalmaktadır. Ayrıca enlem veya deniz seviyesine bağlı olarak farklı coğrafik ve meteorolojik olayların etkisiyle dünya yüzeyinde büyük oranda değişikliğe uğramaktadır. KGI olarak bilinen bu değer güneş tarafından yeryüzüne ulaşan doğrudan ve yayılı ışınım şiddetlerinin toplamıdır ve dünyanın farklı noktalarına farklı şiddette düşer. Ölçülen KGI verisine göre yapılan tasarımların yatırım açısından uygunluğu değerlendirilebilmektedir. Bu değer sadece PV-tabanlı santrallerin tasarımlarında gerekli olmayıp aynı zamanda güneş enerjisi uygulamaları ile alakalı bilimsel ve mühendislik birçok çalışmada da en önemli girdi parametresidir. Bu sebeple seçilen özel bir coğrafik bölge içinde güneş ışınımlarının farklı türlerini ölçen Piyranometre veya Pirheliometre gibi cihazların kurulması ile uzun süreli ölçüm verileri elde etmek en doğru yöntemdir. KGI'nı ölçülmesinde yaygın kullanılan Piyranometre cihazının yapısı Şekil 1.2'de verilmiştir. Cihaz üzerinde siyah yüzeye düşen ışınım değeri arttığında, sıcaklık değeri artacaktır. Bu sıcaklık termocouple modülü tarafından bir voltaj değerine dönüştürülür. Sıcaklık farkı siyah yüzey tarafından emilen ışınım değeriyle ilgili olduğundan, sistem çıkış voltajı ışınımla doğru orantılı olarak azalıp

artmaktadır. Piyranometre kullanarak hem dağınık güneş ışınımı hem de toplam ışınımı ölçülebilmektedir (Coulson, 2012). Ancak kullanılan bu cihazlar pahalı ve hassas olduğu için herhangi bir yere kurulması ve işletilmesi, zaman, maliyet ve uzmanlık gerektirdiğinden, bu çoğunlukla mümkün olamamaktadır. Birçok ülkede olduğu gibi ülkemizde de KGI değerleri her yerde sağlıklı bir şekilde ölçülememektedir. Ülkemizde güneş ışınım verileri, T.C. Tarım ve Orman Bakanlığına bağlı Meteoroloji Genel Müdürlüğünün bu amaçla belirlenen bazı ölçüm istasyonlarından, farklı devlet kurumlarından ve bazı üniversitelerin araştırma amacıyla kurdukları ölçüm istasyonlarından sağlanabilmektedir (Yıldırım, 2016).



Şekil 1.2. Piyranometre cihazının temel yapısı (Raja, 2019)

1.2. Güneş Işınımı Tahmin Yöntemleri

Güneş ışınım değerlerinin ölçümünde yaşanan problemler, cihazların yüksek kalibrasyon ve uzmanlık gerektirmeleri, ölçüm sürecinin oldukça maliyetli olması, her noktada ölçüm yapılamaması, ölçüm veri setinde hatalı ve eksik verilerin bulunması sebebiyle dünya genelinde küresel güneş ışınımını değerini hesaplayabilmek amacıyla Deneysel, İstatistiksel, Yapay Zekâ (YZ) ve birden fazla YZ tekniğinin bir arada kullanıldığı Hibrit tabanlı tahmin yöntemleri geliştirilmiştir (Ayodele ve ark., 2016). Bu yöntemler içerisinde son zamanlarda en sık kullanılan ve başarı performansı yüksek yöntem YZ temelli yöntemlerdir.

1.2.1. Deneysel yöntemler

Temeli matematiksel korelasyonlara dayanan doğrusal veya doğrusal olmayan yapıda eşitlikler kullanarak bir yerin güneş radyasyonunu tahmin yöntemleridir. Temeli

Angström'un regresyon modeline dayanmaktadır ve bu model güneş ışınım tahmini için geliştirilen ilk doğrusal modeldir (Angstrom, 1924). Modelin genel formülü denklem 1.1'de verilmiştir. Bu formülde *H*, yatay bir yüzeye düşen aylık ortalama günlük küresel radyasyonunu, H_0 tamamen açık gökyüzü koşullarında aylık ortalama günlük küresel radyasyonunu, *S*, gün içinde alınan aylık ortalama günlük güneşlenme süresi miktarını, S_0, gün içinde alınabilecek maksimum aylık ortalama günlük güneşlenme süresi miktarını ve *a, b* regresyon katsayılarını temsil eder. İlerleyen süreçte Prescott, Angström'ün formülündeki H_0 parametresinde karşılaşılan zorlukların üstesinde gelmek için bu parametreyi, yatay bir yüzey üzerine düşen dünya dışı radyasyon olarak hesaplayıp, değiştirerek, formülün kullanılabilirliğini arttırmıştır (Prescott, 1940). Bu sebeple formül yaygın olarak Angström-Prescott denklemi olarak bilinmektedir. Sonraki yıllarda formül üzerinde yapılan çeşitli modifikasyonlar neticesinde geliştirilen modellere farklı boyutlar kazandırmak hedeflenmiştir. Tüm modellerde temel girdi parametresi güneşlenme süresidir. Farklı yazarlar enlem, boylam, bağıl nem, hava sıcaklığı, yağış, bulutluluk gibi meteorolojik ve coğrafik girdi verilerinden iki veya daha fazlasını kullanarak hedef bölgeye özgü modellerin, farklı tiplerini geliştirmişlerdir. Bu hesaplamalar sonucunda çeşitli türden ampirik korelasyonlar yardımıyla herhangi bir yerin matematiksel bir formül ile ifade edilmesi/modellenmesi sağlanmış olmaktadır (Besharat ve ark., 2013).

$$\frac{H}{H_0} = a + b\left(\frac{S}{S_0}\right) \tag{1.1}$$

Geliştirilen bu modeller 4 farklı grupta sınıflandırılabilirler.

i) Güneşlenme süresi temelli modeller
ii) Bulut temelli modeller
iii) Sıcaklık temelli modeller
iv) İki veya daha fazla meteorolojik verilerle geliştirilen modeller

Literatüre bakıldığında bu alanlarda yapılan birçok çalışmaya rastlanmaktadır. Bunlardan en yaygın kullanılanı güneşlenme süresi temelli modellerdir (Besharat ve ark., 2013).

1.2.2. İstatistiksel yöntemler

İstatistiksel yöntemler, geçmiş gözlemleri, gözlemleri etkileyen faktörleri, farklı yöntem ve tekniklerle matematiksel olarak modelleyerek, geleceğe yönelik tahminde bulunulmasını sağlayan yöntemler bütünüdür. Bu yöntemlerle, geçmişte ölçülen veriler kullanılarak geleceğe yönelik projeksiyonlar yapılır. Aslında istatistiksel yöntemler, YZ alanına giren Yapay Sinir Ağları (YSA), Destek Vektör Makinası (DVM) gibi birçok algoritmayı da kapsayan oldukça geniş bir alan olarak bilinmektedir. Ancak, bu bölümde güneş ışınım tahmininde sıklıkla kullanılan ve YZ'nın dışında kalan Box-Jenkins tabanlı yöntemler bu sınıfa dahil edilmiştir. Box-Jenkins yöntemi, otoregresif entegre (I) hareketli ortalama (ARIMA) modelleri olarak bilinmektedir. Bu yöntem tek değişkenli ve kısa dönemli tahminlerde tercih edilen bir modellerdir. Özellikle zaman serisi verilerinin analizinde kullanılan bu yöntemde, seçilen zaman aralığında, gözlemlenen veri serisinin yapısını oluşturan stokastik (olasılıklı) sürecin modellenmesi hedeflenmektedir. Kullanılan zaman serisinin, eşit zaman aralıklarıyla elde edilen veri setinden oluşan kesikli ve durağan bir seri olması beklenmektedir. Box-Jenkins yönteminin amacı, en az sayıda değişken ile en uygun modelleri elde etmektir (Box ve Jenkins, 1970). Zamansal veriler, gözlemlenen herhangi bir alana ait gözlem sonuçlarının zaman parametresine göre sıralanmasıyla meydana getirilen veri dizileridir. Geliştirilen tahmin modelleri, yapısı gereği zamansal verilerden etkilenirler ve bu veriler genel olarak stokastiktir. Bu nedenle, serinin gelecek dönemde alması muhtemel değerleri, geçmiş değerleri tarafından kısmen tanımlanabilir. Geliştirilen modellerde, birbirine zamanca yakın veriler, genellikle uzak olanlara göre çok daha fazla benzerlikler gösterirler ve bu benzerlikler bir bağımlılık yaratmaktadır. Bu nedenle zaman serileri analizi, bu bağımlılığın analizine yönelik tekniklerle ilişkilidir (Erdoğan, 2006).

Güneş tabanlı zaman serisi verileri için geliştirilen tahmin modellerinden ilki Otoregresif Hareketli Ortalama (ARMA) modelidir. ARMA, yalnızca durağanlık olması durumunda, yani ortalama ve varyans gibi istatistiksel parametrelerin zaman içinde sabit kaldığı zaman serisi verileri için kullanılabilir bir modeldir. Bu sebeple kullanım alanı sınırlıdır. ARMA, AR (Otoregresif) ve MA (Hareketli Ortalama) gibi iki ana yönteme bağlı olarak geliştirilmiştir. AR, zaman içinde verilerin değişmeyeceği teoremine dayanmaktadır. Bir sistemden kaç birim çıkış var ise, çıkış sayısı kadar sisteme giriş olmalıdır yaklaşımı benimsenmiştir. AR modelinde, m, otomatik gerileyen işlem sırası,

AR(m), denklem 1.2'de verilmiştir. MA modeli ise, serinin gecikmeli hata teriminin, şimdiki hata terimini etkileme durumunu ifade eden bir yaklaşımdır. Bu yaklaşım, denklem 1.3'te MA modelinde, n, hareketli ortalama işlem sırası MA(n) olarak gösterilmiştir. ARMA yöntemi ise geliştirilen modellerde az sayıda parametre ile daha esmek bir yapı oluşturmak adına hem otoregresif hem de hareketli ortalama parametrelerinin dahil edilmesi fikrine dayanmaktadır. Doğrusal bir fonksiyon temelinde, m ve n parametrelerine göre geliştirilen modelin formülü ARMA(m,n) denklem 1.4'te verilmiştir. Durağan olmayan zaman serilerinde, ARMA modelinin genelleştirilmiş şekli olan ARIMA modeli kullanılmaktadır. ARIMA modeli ise Box-Jenkins tarafından iki modelin birleştirilmesi ile oluşan sayısal bir modeldir ve tek veya çok değişkenli yapıda modeller geliştirilebilmektedir. Bu model, güneş radyasyonu zaman serileri tahmin etmek için en uygun model olarak kabul edilmektedir. Bu tipte geliştirilen modellerde amaç, en uygun ve en az değişken ile doğrusal modeli hesaplamaktır. ARIMA (m,k,n) modelinin formülü denklem 1.5'te verilmiştir (Sobri ve ark., 2018; Shadab ve ark., 2019). Bahsi geçen bu yöntemler, 1970'lerden beri küresel güneş ışınımının tahmin edilmesinde kullanılan en eski yöntemler olarak güncelliğini korumaktadır (Inman ve ark., 2013).

$$G_t = \sum_{i=1}^{m} \phi_i G_{t-i} + \omega_t \tag{1.2}$$

$$G_t = \omega_t + \sum_{j=1}^{n} \alpha_j \omega_{t-j} \tag{1.3}$$

$$G_t = \sum_{i=1}^{m} \phi_i G_{t-i} + \sum_{j=1}^{n} \alpha_j \omega_{t-1} + \omega_t \tag{1.4}$$

$$G_t = \sum_{i=1}^{m} \phi_i S^k G_{t-i} + \sum_{j=1}^{n} \alpha_j \omega_{t-1} + \omega_t \tag{1.5}$$

Bahsi geçen formüllerde, G_t zaman serisi değerini, $\phi\,(\phi_1+\phi_2+\phi_3+...+\phi_m)$ ve $\alpha\,(\alpha_1+\alpha_2+\alpha_3+...+\alpha_m)$ AR ve MA model katsayısı vektörlerini, ω_t hata değerlerini, m otoregresif eleman sayısını, n ise gecikmeli tahmin hatalarının sayısını, S AR(m) operatörünü, k durağan veri sayısı ifade eder. Literatürde ARIMA yönteminin, güneş ışınım tahmin çalışmalarında sıklıkla kısa vadeli saatlik veya günlük radyasyon verileri kullanılarak, tahmin modellerinin geliştirildiği görülmektedir. Uzun süreli ölçülen güneş

ışınım verilerine dayalı çalışmalarda çok sık kullanılan bir yöntem değildir (Shadab ve ark., 2019).

1.2.3. Yapay zekâ yöntemleri

YZ, bilgisayar temelli makinelerin karmaşık problemlere, insanlara benzer çözümler üretmesini, akıllı makineleri ve özellikle de akıllı bilgisayar programlarının yapılmasını amaç edinen bir bilim dalıdır ve aynı zamanda YZ, tamamen yapay araçlar ile geliştirilen ve insan gibi davranışlar ve hareketler sergileyebilen program ve makinelerin meydana getirilmesini sağlayan mühendislik alanı olarak bilinmektedir. YZ teknolojileri, her çeşit makinenin deneyimlerden öğrenmesini, sistemin karşılaştığı yeni girdilere uyum sağlamasını ve insan temelli görevleri yerine getirmesini mümkün kılan araçları kapsamaktadır. Bilgisayar biliminde YZ, akıllı ajanlar olarak tabir edilen nesnelerin çalışması sonucunda, çevresini algılayan ve bir hedef için başarı performansını oldukça üst seviyelere çıkarmayı amaç edinen eylemleri gerçekleştirebilen bilgisayar tabanlı sistemler olarak tanımlanmaktadır. Süreç sonunda geliştirilen sistemlerden, algılama, öğrenme, düşünme, fikir yürütme, problem çözme, iletişim sağlama, çıkarım yapma, karar verme, nesneleri hareket ettirme ve kullanma becerisi gibi önemli bilişsel fonksiyonların veya otonom davranışların gerçekleştirilmesi beklenmektedir. YZ algoritmalarında, istatistik, hesaplamalı zekâ ve klasik sembolik zekâ gibi yaklaşımlarından, ayrıca matematik, psikoloji, dilbilimi, felsefe, sinir bilimi gibi birçok farklı bilim dallarından yararlanılmaktadır. Son yıllarda oldukça popüler olan bu kavram Artificial Intelligence (AI) olarak bilinmekte ve bilgisayar temelli her teknolojide oldukça fazla yararlanılmaktadır (Ongsulee, 2017).

YZ'nın bu kadar yaygın olmasının ana sebebi, insanın düşünme ve problem çözme becerilerini keşfederek, bunun yapay yönergeleri geliştirme çabası ve makinalara aktarma yöntemleri olmuştur. Özellikle son yıllarda hızla artan veri miktarları, problemlerin karmaşıklığı ve bilgisayarların hesaplama kabiliyetlerinin artması, yeni makine öğrenme teknolojileri ve uygulamalarının geliştirilmesinde, üniversitedeki araştırmacıları, endüstri ve teknoloji şirketlerini, bu alana ilgi duyan yatırımcıları ve ülkeleri oldukça büyük paralar harcayarak yoğun ilgi göstermeye zorlamaktadırlar. Sadece tahmin etme yeteneklerinin geliştirilmesinde değil, aynı zamanda, daha sofistike algoritmalar üreterek, kişilerin, toplumların ve işletmelerin geleceğine yön verecek insani kararların alınmasını sağlayacak otonom modellerin geliştirilmesi

hedeflenmektedir. Günümüzde tıptan, mühendisliğe, bankacılık ve finans hizmetlerinden, turizme kadar hemen hemen her alanda çalışabilen uygulamalar geliştirilmiştir. Dil çeviri sistemleri, hava alanı kontrol sistemleri, robotik sistemler, tıbbi teşhis sistemleri gibi sistemler bu alanda geliştirilen önemli uygulamalardandır (IEC, 2018).

YZ, ilk olarak 1956 yılındaki Dartmouth Konferanslarında bir grup bilgisayar bilimcisinin bu kavramı ortaya atmasından bu yana gelişimini sürdüren bir alan olarak bilinmektedir. 1958 yılında Frank Rosenblatt tarafından ilk sinir ağı (perceptron) algoritmasının keşfi, YZ'nın gelişimindeki en önemli adım sayılmaktadır. Üzerinde sürekli araştırma gerçekleştirilen YZ, gerçek bir uygulama üzerinde ilk kez 1967 yılında en yakın komşu makine öğrenme algoritmasının geliştirilmesiyle uygulanmış ve bu süre zarfında oldukça sınırlı bir uygulama alanına sahip olmuştur. Özellikle 1980'lere kadar geçen süre sinir ağları algoritmalarının gelişim süreci olarak bilinen YZ'nın temelinin inşa süreci olarak adlandırılmaktadır. 2000'li yıllara kadar YSA oldukça popüler olmuş ve hem birçok sinir ağı mimarisi geliştirilmiş hem de birçok alanda uygulamalar yapılmıştır. YSA'nın bu kadar yaygınlaşmasında, 1986 yılında geri yayılım algoritmasının geliştirilmesi önemli katkılar sunmuştur. Bununla birlikte bu dönemde uzman sistemler, DVM algoritması ve matematiksel modelleme alanında gerçekleştirilen önemli keşifler YZ tekniklerinin kullanımını ve iyileştirmesini arttırmıştır. 1980'lerden başlayarak içinde bulunduğumuz 21. yüzyılın başlarına kadar geçen sürede ML algoritmalarının yoğun kullanımı ve birbirinden farklı alanlarda yapılan on binlerce farklı uygulama YZ'yı oldukça popüler hale getirmiştir. Özellikle son on yıldır derin öğrenme sinir ağları algoritmalarının geliştirilmesi ve iyileştirilmesi ile Derin Öğrenme (DL) tabanlı yapılan uygulamaların artması, YZ'nın yeniden doğuşu olarak adlandırılan bir sürecin başlamasına yol açmıştır. Ayrıca son zamanlarda birden fazla ML algoritmasından veya farklı yazılım ve YZ yönteminden faydalanılarak hibrit yöntemler olarak adlandırılan tahmin modelleri de geliştirilmiştir. Bu modellerde, genellikle, bir arada çalışan yöntemlerden biri ile tahmin modelleri geliştirilirken, diğer yöntem ile veriler veya algoritma üzerinde optimizasyon işlemleri yapılmaktadır. Hibrit olarak geliştirilen modellerde asıl amaç tahmin kesinliğini arttırmak ve birden fazla YZ tekniğinin kolektif bir biçimde çalışmasını sağlamaktır (Casti, 2004; IEC, 2018).

1.2.3.1. Makine öğrenme yöntemleri

YZ, temelde sembolik ve ML yöntemleri üzerine inşa edilmiştir. İlk geliştirilen uygulamalar sembolik yöntemler kullanılarak oluşturulmasına karşın, ML tekniklerinin geliştirilmesi ile dikkatler bu alana odaklanmıştır. YZ, günümüzde makineleri daha akıllı hale getirmeye hedefleyen ve bu amaçla kullanılan birçok tekniğin ML algoritmalarını içeriyor olması nedeniyle ML'ne daha yakın bir disiplin alanı olarak bilinmektedir.

Artur Samuel, 1959 yılında, ML'ni bilgisayarlara açıkça programlanmadan öğrenme yeteneği veren geniş bir YZ alt çalışma alanı olarak tanımlamıştır. O günden bu güne ML, çeşitli türden geçmiş veriler veya bilgiler arasındaki ilişkilerden yararlanarak öğrenmek, deneyim ve eğitim yolu ile tahminler yapmak, sistem performansını iyileştirmek ve kararlar almak amacıyla farklı türlerden algoritmaların incelenmesini ve geliştirilmesini sağlayan hesaplama yöntemleri olarak tanımlanmıştır (Awad ve Khanna, 2015). ML'nin temel amacı, makinaların önceden bilmediği gelecekteki olayların veya senaryoların tahmin edilebilmesini sağlamaktır. Bu sebeple, lojistik regresyondan, karar ağaçlarına, YSA'dan DL'ye kadar oldukça geniş teknikleri kapsayan ve hemen hemen tüm çalışma alanlarının verilerine uygun tahmin modelleri geliştirilmiştir.

ML modellerinin birçok alanda olduğu gibi enerji sistemlerinin modellenmesi, tasarımı ve tahmininde 2000 yıların başından beri gittikçe artan oranlarda kullanılması, farklı ML modellerinin geliştirilmesini ve uygulanmasını önemli oranlarda arttırmıştır. Bu artışta gittikçe artan veri miktarları ve bu verilerin işlenmesinde kullanılan farklı yöntem ve tekniklerin performanslarının düşük kalması yatmaktadır. Ayrıca ML yöntemlerinin tahmin doğruluğu ve hızlarının oldukça yüksek seviyelere ulaşması yaygınlığının artmasındaki en önemli parametrelerden birisi olmuştur. Özellikle, enerji alanında farklı ML algoritmaları kullanılarak yapılan, Web of Science, Scopus gibi önemli veri tabanlarında yayınlanan araştırmalar, 2019 yılının bir önceki yılına göre yaklaşık %45'lik bir artış oranı yakalamış ve çalışmaların çoğu elektrik enerjisi ve yenilenebilir enerji alanlarına odaklanmıştır (Mosavi ve ark., 2019).

ML, güneş tabanlı çalışmalarda oldukça fazla sayıda uygulaması olan bir tahmin yöntemidir ve güneş ışınımını tahmin edebilmek amacıyla dünyanın farklı coğrafik alanlarında, farklı türden algoritmalar kullanılarak birçok araştırma yapılmıştır. Yapılan çalışmalarda, YSA, DVM / DVR, K-NN, Doğrusal / Doğrusal olmayan regresyon, M5, Rastgele Ormanlar gibi algoritmalar sıklıkla kullanılmaktadır (Voyant ve ark., 2017). Bu araştırmalarda, bir ülkenin belirli bir coğrafik bölgesi ya da ülke içerisindeki farklı

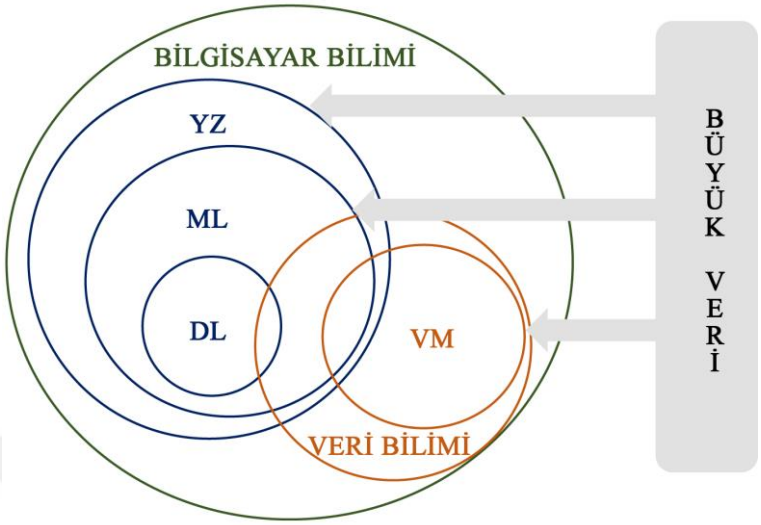
coğrafik yerler seçilerek tahmin modelleri geliştirilmiştir. Modellemeye başlarken, ilk olarak yatay bir yüzeye düşen saatlik, günlük veya aylık ortalama KGI değerlerinden hangisinin kullanılacağına karar verilmektedir. Ardından, çalışmanın türüne göre girdi verileri toplandıktan sonra bir, birden fazla veya hibrit ML algoritmaları kullanılarak hedef bölgenin güneş ışınım değerleri tahmin edilmektedir. Geliştirilen tahmin modellerinde KGI tahmin problemlerine, tek bir ML algoritmasının fonksiyonel yapısı ve mimarisi üzerinde değişiklikler yapılarak, birden fazla algoritmayı karşılaştırarak veya iki veya daha fazla YZ yöntemini bir arada çalıştırarak farklı çözümler bulunmaya çalışılmıştır (Muneer ve ark., 2015).

ML ile tahmin yapılırken amaca uygun verinin seçilmesi, düzenlenmesi ve bunlar üzerinde çeşitli işlemlerin yapılması en önemli konuların başında gelmektedir. Çünkü ML algoritmaları veri odaklı yapılar olarak bilinmekte ve çalışma mantığı gereği verinin türü, yapısı, uzunluğu ve büyüklüğünden doğrudan etkilenebilmektedir. Dolayısıyla ML algoritmalarının durağan bir yapısı yoktur ve her algoritma her veri setinde aynı yüksek performansı gösterememektedir. Algoritma, verileri analiz etmekte, öğrenme planını belirlemekte ve ardından sonuca ulaşacak prosesleri yerine getirebilmektedir. Temelde, verilerden öğrenmektedir ve bu süreçte verilerden anlamsal bilgiler de keşfedilebilmektedir. Bu sebepten, veriler açıkça ML süreçlerinde birincil derecede önemli bir rol oynamaktadır ve öğrenme sürecinin nasıl gerçekleşeceğini belirleyen esas olarak verilerin yapısı olmaktadır. Güneş tabanlı çalışmalarda, ışınıma doğrudan ve dolaylı yoldan etki eden veya edebilecek birçok coğrafik (enlem, boylam, yükseklik gibi), meteorolojik (sıcaklık, basınç, rüzgâr hızı, güneşlenme süresi gibi) ve kategorik (yıl, ay, saat gibi) veri türü bulunduğundan geliştirilecek modellerde yararlanılacak değişkenlerin sayısı ve miktarı günden güne önemli oranda artmaktadır. Bu veriler saatlik ortalama, günlük ortalama ve aylık ortalama gibi değişik formlarda çalışmalara dahil edilerek kullanılmaktadır. Bu nedenle tahmin modellerinde oldukça fazla ve farklı türdeki veri deseni ile çalışılmaktadır. Seçilen veri seti üzerinde hiçbir işlem yapılmadan modelleme çalışmalarında kullanılması başarı performansı üzerinde olumsuz bir etkiye sahip olduğu bilinen bir gerçekliktir. Bu nedenle çalışmaya başlamadan önce eldeki ham veri seti üzerinde takım işlemlerin yapılması tahmin performansını önemli ölçüde arttırmaktadır (Long ve ark., 2014; Makridakis ve ark., 2018).

1.3. Veri Madenciliği

Günümüzde veri miktarında meydana gelen büyük artışlar yararlı veri desenlerinin ortaya çıkarılmasını kolaylaştıracak yöntemleri kullanmayı zorunlu kılmaktadır. Sıklıkla kullanılan klasik yöntemlerle var olan verilerden anlamlı bilgilerin elde edilmesinin çok kolay olamayacağı anlaşılmıştır. Bu sebeple veri madenciliği olarak bilinen yöntem ve tekniklerden faydalanılarak veriler üzerinde birtakım prosesler yapılmaktadır. Veriler arasında bilgiye ulaşma, bilgiyi analiz etme, bilgiden anlam çıkarma ve tahmin yürütme bilimine genel olarak VM denir ve kavram olarak ilk defa 1960'lı yıllarda ortaya çıkmıştır. Bu yılarda veri taraması veya veri yakalanması gibi isimlerle anılmıştır. İlişkisel veri tabanı yönetim sistemleri 1970'lerde, 1980'lerde ise veri tabanı yönetim sistemlerinin birçok alanda yaygın hale gelerek kullanımı artmıştır. Karmaşık veri desenlerinden yararlı ve faydalı bilgilerin çıkartılması amacıyla, ham verilerden bilgi keşfi süreci olarak tanımlanan ve Knowledge Discovery of Data (KDD) olarak bilinen süreç, özel bir alan olarak 1989 yılından beri uygulanmaktadır (Fayyad ve ark., 1996). Bilgisayar bilimi içerisinde kendine yer edinmiş özel bir veri bilimi alanıdır ve YZ'nın çalışma alanları için bir temel görevi görür. Şekil 1.3'te veri madenciliğinin bilgisayar bilimi içindeki diğer alanlar ile ilişkisel durumu gösterilmiştir. VM ve ML yöntemleri temelde aynı algoritmaları kullanırlar. Ancak, genel olarak, VM daha önce bilinmeyen özelliklerin keşfine odaklanırken, ML algoritmaları bilinen özelliklere dayanarak tahminde bulunmaya odaklanır (Witten ve ark., 2011). VM, hemen hemen tüm alanlarda gizli bilgilerin açığa çıkarılabilmesini ve gelecekteki eğilimlerin tahmin edilebilmesini sağlamada kullanılır ve birçok alanda uygulaması olan bir tekniktir. Özellikle eğitim, pazarlama, sigortacılık, üretim, bilim ve mühendislik, bankacılık ve borsa, astronomi, biyoloji, internet, tıp, telekomünikasyon gibi alanlarda yaygın olarak kullanılmaktadır. Enerji alanında yeni olsa da binaların enerji tüketimi ve taleplerinin modellenmesi, HVAC (ısıtma, soğutma, havalandırma ve iklimlendirme) olarak adlandırılan sistemlerin enerji verimliliğinin optimizasyonu ve modellenmesi, coğrafi bir bölgenin enerji talepleri ve tüketiminin modellenmesi, elektrik fiyatı tahmini, soğutucu akışkanlarının özelliklerinin tahmini, bazı sistemlerin enerji verimliliği tahmini, rüzgar türbini enerji gücü optimizasyonu ve verimliliği tahmini, çeşitli türdeki kazanların enerji verimliliği tahmini gibi enerji alanına giren konularda yapılan çalışmalara rastlanmaktadır. İçerisinde makine öğrenimi, sinir ağları, istatistik, veri tabanı teknolojileri, görüntüleme gibi birçok disiplini barındırırlar. Genel olarak tahmin

edici ve tanımlayıcı VM modelleri kullanarak veriler üzerinde işlemler yapılmaktadır (Selbaş ve ark., 2011).

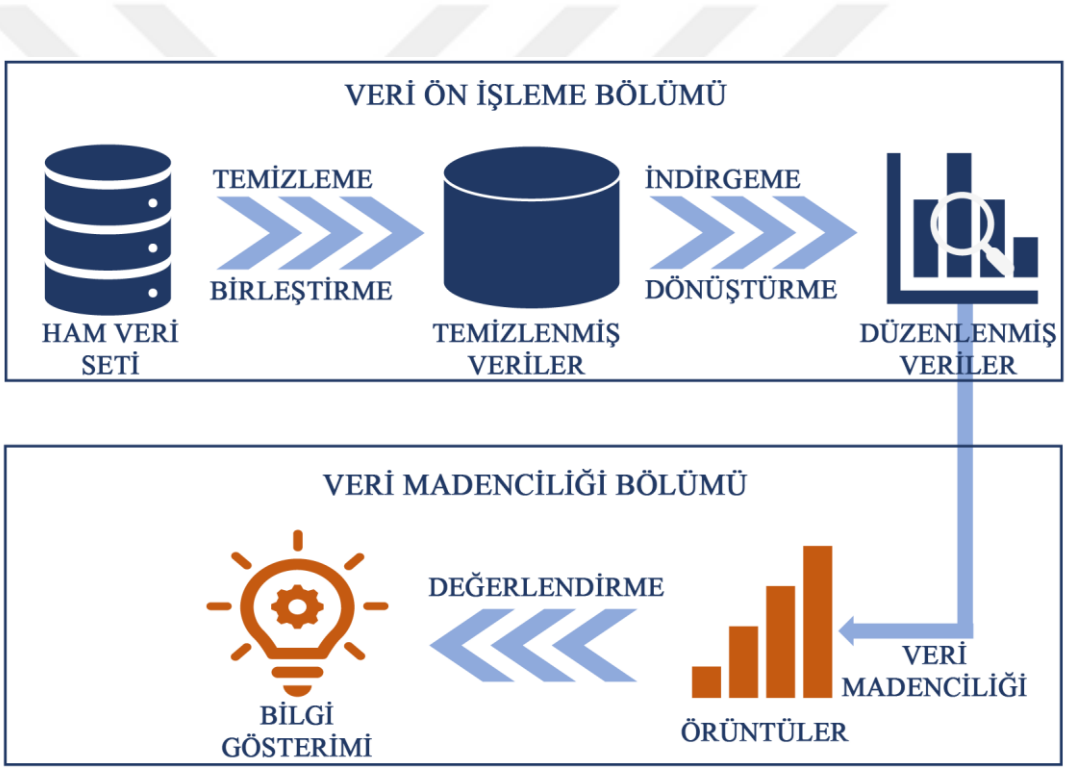


Şekil 1.3. Veri Madenciliğinin bilgisayar bilimi alanları ile ilişkisel yapısı

VM ile YZ tabanlı uygulamaların ortak parametresi kullanılan veri setidir. Geliştirilen uygulamaların birçoğunda temel amaç veri seti ile uyumlu en başarılı tahmin modelini veya modellerini geliştirebilmektir. Ancak veri setinin yapısı ve uzunluğu istenilen seviyede tahmin doğruluğuna ulaşmayı engelleyebilmektedir. Veri seti içindeki değişkenler ne kadar fazla olur ise modelin hedef çıktısı ile ilişkilendirmede sorunlar yaşanabilmektedir. Bu sebeple, eldeki verilerin hepsini kullanmak her zaman çıktı üzerinde beklenilen üst düzey performans seviyesine ulaşmada sorunlar yaratabilmektedir. Bir tahmin sonucuna ulaşılır, fakat "Daha iyi bir performans elde edebilir miydim?" diye akıllarda bir soru işareti kalmaktadır. Bunun önüne geçmek için bazı çalışmalarda girdi verilerinin farklı yöntem ve teknikler kullanılarak veya deneme-yanılma yoluyla çeşitli kombinasyonları alınarak tahmin modelleri geliştirilmektedir. Son zamanlarda ML algoritmalarında bir model geliştirmeden önce veri üzerinde çeşitli işlemler yapılarak amaca uygun analiz edilmesi hedefe ulaşmada başarı performansını arttırmakta ve tahmin sürecini kısaltmaktadır.

VM, veri bilimi içerisinde önemli bir yere sahip olmasına rağmen, büyük ve karmaşık veri setlerinden amaca uygun bilgilerin kolay yoldan alabileceği basit bir yöntem değildir. Genellikle karmaşıktır ve bilgisayar bilimi içerisindeki birçok yöntem ve teknikten faydalanılarak uygulanan madencilik süreçleridir. Bu süreçlerin kendi

içerisinde bir hiyerarşisi ve işlem adımları bulunmaktadır. Veri madenciliği süreçleri, veri ön işleme (hazırlama) ve VM olmak üzere iki ana bölüme ayrılmaktadır. Şekil 1.4'te veri madenciliği süreçlerinin bölümleri ve işlem prosesleri gösterilmiştir. Veri ön işleme bölümü, veri temizleme, veri birleştirme, veri indirgeme ve veri dönüştürme adımlarından oluşurken, VM bölümü ise, veri madenciliği, örüntü değerlendirmesi ve bilgi gösterimi adımlarından oluşmaktadır. Veri madenciliği adımında düzenlenmiş verilerden gizli veya bilinmeyen örüntüleri çıkarmak amacıyla sınıflandırma, tahmin, kümeleme, zaman serisi analizi gibi birçok yöntemin uygulandığı teknikler yerine getirilmektedir. Veri madenciliği süreci veri ön işlemenin temizleme adımından başlar ve VM'nin bilgi gösterimi adımında son bulan tekniklerini kapsayan işlem basamaklarından oluşmaktadır (Famili ve ark., 1997; Susanto ve Meiryani, 2019).



Şekil 1.4. Veri madenciliği süreçleri işlem prosesleri (WideSkills, 2015)

Veri temizleme adımı veri madenciliği sürecinin ilk adımı olarak bilinir ve ham verilerdeki eksik ve hatalı (gürültülü) verilerin düzeltilmesi ve tutarsızlıkların giderilmesi amacıyla uygulanır. Süreç içerisindeki bir başka adım olan veri birleştirme tekniğinde farklı veri kaynaklarından alınan veriler tek bir veri kaynağı yapısında birleştirir. Farklı tiplerdeki veri kaynaklarının yapısına göre çok zor ve karmaşık bir süreç olabilmektedir. Veri dönüştürme, ham verilerin bir formattan, başka bir veri

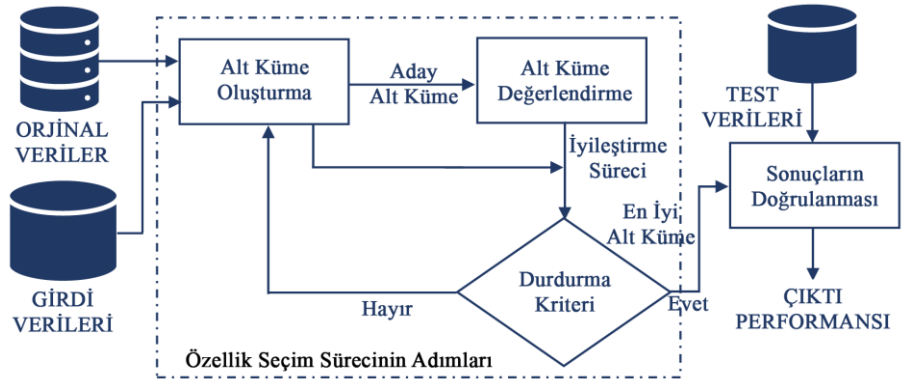
formatına veya yapısına ölçeklendirme, normalleştirme veya ayrıştırma işleminin veya işlemlerinin uygulanması sürecidir. Bu sayede veriler VM veya ML algoritmaların doğrudan uygulanabileceği bir yapıya dönüştürüldüğünden tahmin performansı üzerinde pozitif bir etki göstermektedirler. Veri indirgeme adımında ise, veri setinden ilgisiz özelliklerin (değişken) elenmesi, ilgili olanların seçilmesi, sıralanması veya boyutunun azaltılması amaçlanır. Veri ön işleme süreçleri olarak bilinen bu dört süreç, insan etkileşimi gerektirdiğinden zor, karmaşık ve zaman alıcı bir yapıdadır. Veri madenciliği uygulamalarında toplam harcanan zamanın yaklaşık %80'i veri ön işleme süreçlerine harcanmaktadır (Gibert ve ark., 2016). Bu sebeple, VM veya farklı bir ML uygulamasında kaliteli ve kesin sonuçlar alınmak isteniyorsa, geliştirilen modellerden ilişkisel ve anlamsal bilgilerin edinilmesi amaçlanıyorsa, veri ön işleme adımlarının dikkatli bir şekilde uygulanması ve yeterli zamanın ayrılması gereklidir.

1.3.1. Amaca uygun özelliklerin seçimi

Veri madenciliği süreçleri içerisinde veri setinden amaca uygun değişkenlerin seçilmesi özellik seçimi veya özellik mühendisliği olarak tanımlanmaktadır. Veri ön işleme sürecinde indirgeme adımında uygulanan ve son zamanlarda model geliştiriciler tarafından dikkat çekilen önemli tekniklerin başında gelmektedir. Yapılan çalışmada ilgili özelliklerin tespiti ve alakasız, gereksiz veya gürültülü verilerin kullanılacak veri tabanından kaldırılması işlemidir. YZ tabanlı uygulamaları hızlandırır, tahmin doğruluğunu iyileştirir ve modelin anlaşılırlığı geliştirmektedir. Hem ML algoritmaları hem VM uygulamalarında geliştirilen modelin eğitim süresi ve performansı veri kümesindeki özelliklerle büyük oranda ilişkilidir. Özellik mühendisliği, eldeki verilere, uygulanan modele ve amaca göre en uygun özelliklerin formüle edilmesi sürecidir.

Özellik seçimi süreci basit prosesleri olan bir uygulama olmamakla birlikte karmaşık teknikler, istatistiksel yaklaşımlar ve farklı ML algoritmaları içeren derin ilkeleri bulunmaktadır. Özellik seçimi bireysel ve altküme değerlendirmesi olarak iki ana yaklaşıma ayrılmaktadır. Bireysel değerlendirme, özelliklerin ağırlıklarının alaka seviyesine göre sıralanması tekniğine dayanırken, alt küme değerlendirmesi aday altkümelerin veri setinden farklı yöntemlerle aranması stratejisi üzerine işlem yapmaktadır. Genel olarak özellik seçimi *alt küme oluşturma veya arama*, *alt küme değerlendirme*, *durdurma kriteri* ve *alt kümenin performansının değerlendirmesi* olmak üzere dört aşamadan oluşmaktadır. Şekil 1.5'te özellik seçim sürecinin çalışma

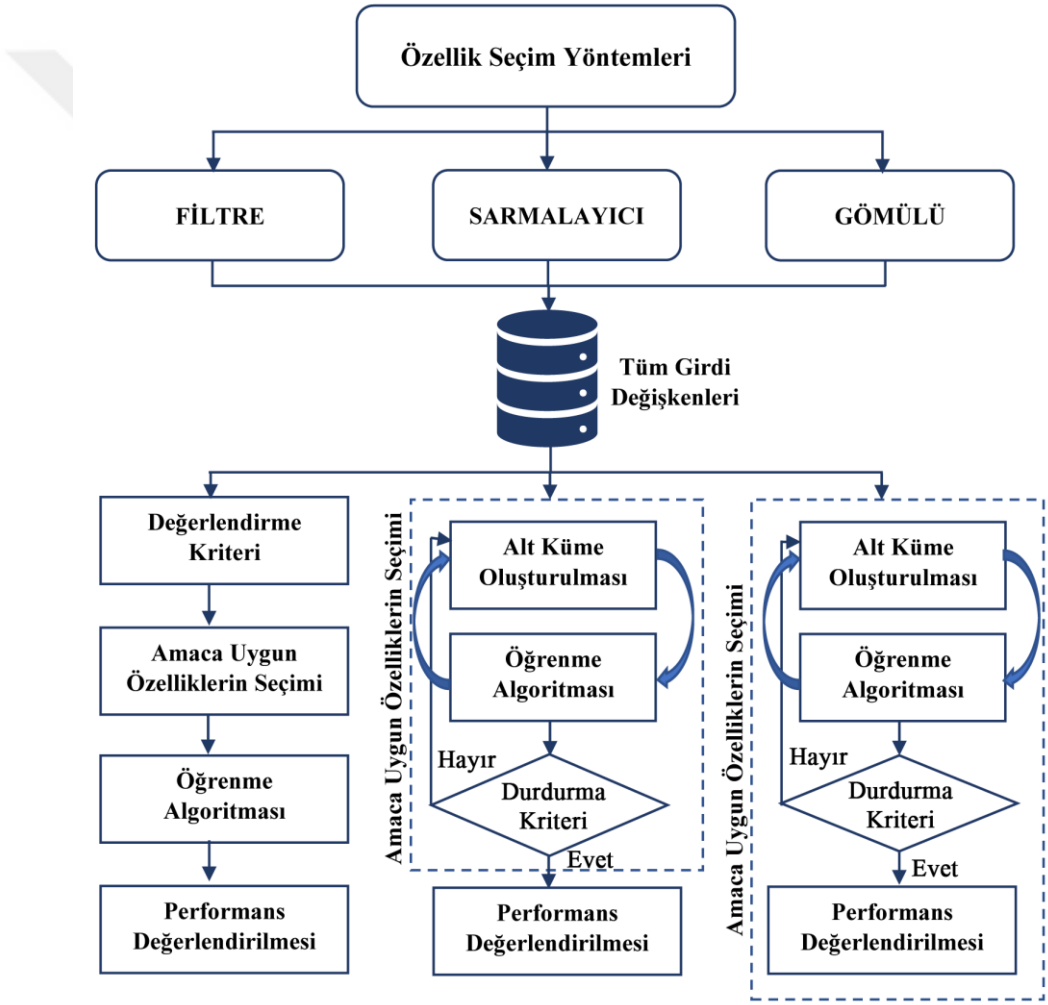
prosedürü akış şeması olarak verilmiştir. İlk üç adım özellik seçim sürecinin ana adımıyken, son adım bu sürece dahil olmayan seçilen özelliklerin değerlendirme sürecini kapsar. Bununla birlikte, özellik seçim süreçlerinde *Sarmalayıcı* (Wrapper), *Filtre* (Filter) ve *Gömülü* (Embedded) olmak üzere üç genel özellik seçim yöntemi kullanılmaktadır. Filtre yönteminde, ML algoritmasından bağımsız olarak girdi verilerinin genel özellikleri kullanılarak en uygun veriler aranmaktadır. Uygulanan yöntem temelde tek değişkenli ve çok değişkenli filtre yöntemleri olmak üzere iki faklı kategoriye ayrılmıştır. Sarmalayıcı yöntemler, özellik seçimini uygulanan tahmin algoritmasının etrafına sarar ve veri setindeki bir dizi değişkenin tekrarlı olarak seçilmesi veya elenmesi için geliştirilen modelin tahmin doğruluğunu kullanmaktadırlar. Bu yöntemde en uygun özellik alt kümesini belirlemek için farklı olası özellik alt kümeleri tanımlanır ve performansları herhangi bir tahmin algoritması ile değerlendirilmektedir. Tahmin modelinin oluşturulmasında seçilen bir dizi özellik ile örnek bir eğitim veri seti oluşturularak eğitilir, ardından geri kalan veriler ile test veri seti oluşturularak geliştirilen model değerlendirilmektedir. Gömülü yöntemler ise seçim süreci tahmin modelinin hesapladığı puanlara dayanmaktadır. Uygulamada girdi verileri ile bir tahmin modeli oluşturulur ve geliştirilen modelin doğruluğu test verileri ile değerlenmektedir. Geliştirilen tahmin modeli, her bir özellik için performans değerlerini hesaplamaktadır. Özellik alt kümesinin seçimi yerine, değişkenlerin her birine önem seviyesine göre performans değerleri atanarak düzenlenmesi yapılmaktadır.



Şekil 1.5. Özellik seçim sürecinin çalışma prosedürü (Huan ve Lei, 2005)

Filtre yöntemleri, çapraz doğrulama performansı yerine tek veya çok değişkenli istatistiksel hesaplanan özelliklerin uygunluğunu ölçerken, sarmalayıcı yöntemler ise tahmin modelinin algoritmasına bağlı olarak elde edilen performansın sonucuna göre

her bir değişkenin etkisini ölçmektedir. Bu sebeple, sarmalayıcı yöntemler, herhangi bir tahmin algoritmasını geliştirilen modelde kullanarak optimize etmesinden dolayı, filtre yöntemine göre daha iyi performans göstereceğini söylemek mümkündür. Ancak, öğrenme adımlarının zaman alması ve çapraz doğrulama süreçleri nedeniyle filtre yöntemlerine kıyasla özellik seçim süreçleri daha zor olmakta ve uzun sürmektedir. Gömülü yöntemler, tahmin modelinin performansını en iyi hale getirmek amacıyla kullanıldığından sarmalayıcı yöntemlere benzer olarak çalışmaktadırlar. Tek fark öğrenme sırasında yapısal bir model oluşturma yöntemini kullanırlar. Şekil 1.6'da filtre, sarmalayıcı ve gömülü özellik seçim yöntemlerinin işlem adımları gösterilmiştir.



Şekil 1.6. Filtre, Sarmalayıcı ve Gömülü özellik seçim yöntemlerinin işlem adımları (Venkatesh ve Anuradha, 2019)

1.4. Çalışmanın Amacı ve Önemi

İçinde bulunduğumuz yüzyılda, küresel ısınma ve buna bağlı sıcaklık artışı, mevsimlerin değişmesi, buzulların erimesi, fırtınalar, kuraklık gibi sebepler, iklim değişikliği gerçeğiyle daha belirgin yüzleşmemize yol açmıştır. İklim değişikliğin en önemli sebepleri arasında fosil yakıtlarının gittikçe artan oranlarda kullanımı yatmaktadır. Bu sebeple, ülkeler enerji politikalarında yenilenebilir enerji kaynaklarının kullanımını ve bu alanda yapılan yatırımları arttırarak fosil yakıtlarının zararlı etkilerini azaltılmayı hedeflemektedirler. Bununla birlikte sınırlı bir enerji kaynağı olan karbon temelli yakıtların yerine rüzgâr, güneş gibi yenilenebilir kaynaklara yönelmek birçok nedenden artık bir zorunluluk olmaktadır. Güneş enerjisi ise yenilenebilir kaynaklar arasında milyonlarca yıldır var olan ve sınırsız bir enerji kaynağı olarak özel bir konuma sahiptir ve son zamanlarda en fazla yatırımın yapıldığı alanların başında gelmektedir. Buradan hareketle, bu tezin konusu olan güneş enerjisi, yenilenebilir enerji kaynağı olarak ülkemiz ve dünyada gittikçe artan önemi göz önüne alınarak belirlenmiştir.

Güneş enerjisinden elektrik üretmek amacıyla yaygın olarak kullanılan santraller büyük oranda PV tabanlıdır. Günümüzde PV maliyetleri önemli oranlarda düşmesine rağmen hala oldukça yüksek bir seviyede bulunmaktadır. Bu sebeple, herhangi bir yere güneş enerji santrali kurulmadan önce yapılacak yatırımın planlanması ve fizibilitesinin hesaplanması gerekmektedir. Hedeflenen bir bölgeye, güneş temelli aktif bir sistem kurulmadan önce ilgili bölgenin güneş ışınım potansiyelinin ölçülmesi, o bölgeye yatırım yapmadan önce, ilk ve temel şartı oluşturur. Güneş enerjisi bir bölgede ölçülen KGI değeri ile doğru orantılıdır. Işınım potansiyeli ne kadar fazla ve sürekli olursa üretilen enerjide o oranda yüksek olacaktır. Ölçülen güneş ışınım verileri yardımıyla ilgili bölgeye özgü yatırım, sistem tasarımı ve planlaması yapılmaktadır. Ancak birçok ülkede KGI verileri sağlıklı bir şekilde ölçülememektedir. Bu sebeple tahmine dayalı yöntemler geliştirilmiştir. Geçmişten bugüne deneysel, istatistiksel ve YZ tabanlı yöntem ve teknikler kullanılarak güneş ışınım değerleri tahmin edilmiş ve edilmektedir. Var olan yöntemlerden en yeni olanı kesinliği sebebiyle YZ'nın bir alt alanı olan ML yöntemleridir. Araştırmacılar KGI tahmin edilmesinde ML yöntemini performans olarak diğer yöntemlerden daha üstün bulduklarından son zamanlarda popüler hale gelmiş ve bu alanda birçok çalışma yapılmıştır. ML içerisindeki YSA, DVR, DVM, K-NN, Karar Ağaçları, M5 Kuralları gibi birçok algoritma ya tek başına ya da bir arada karşılaştırmalı olarak bu çalışmalarda kullanılmıştır. Ayrıca yapılan çalışmalarda bir ülkenin belirli bir coğrafik bölgesi ya da ülke içerisindeki farklı coğrafik yerler seçilerek tahmin modelleri geliştirilmiştir. Geliştirilen tahmin modellerinde yatay bir yüzeye

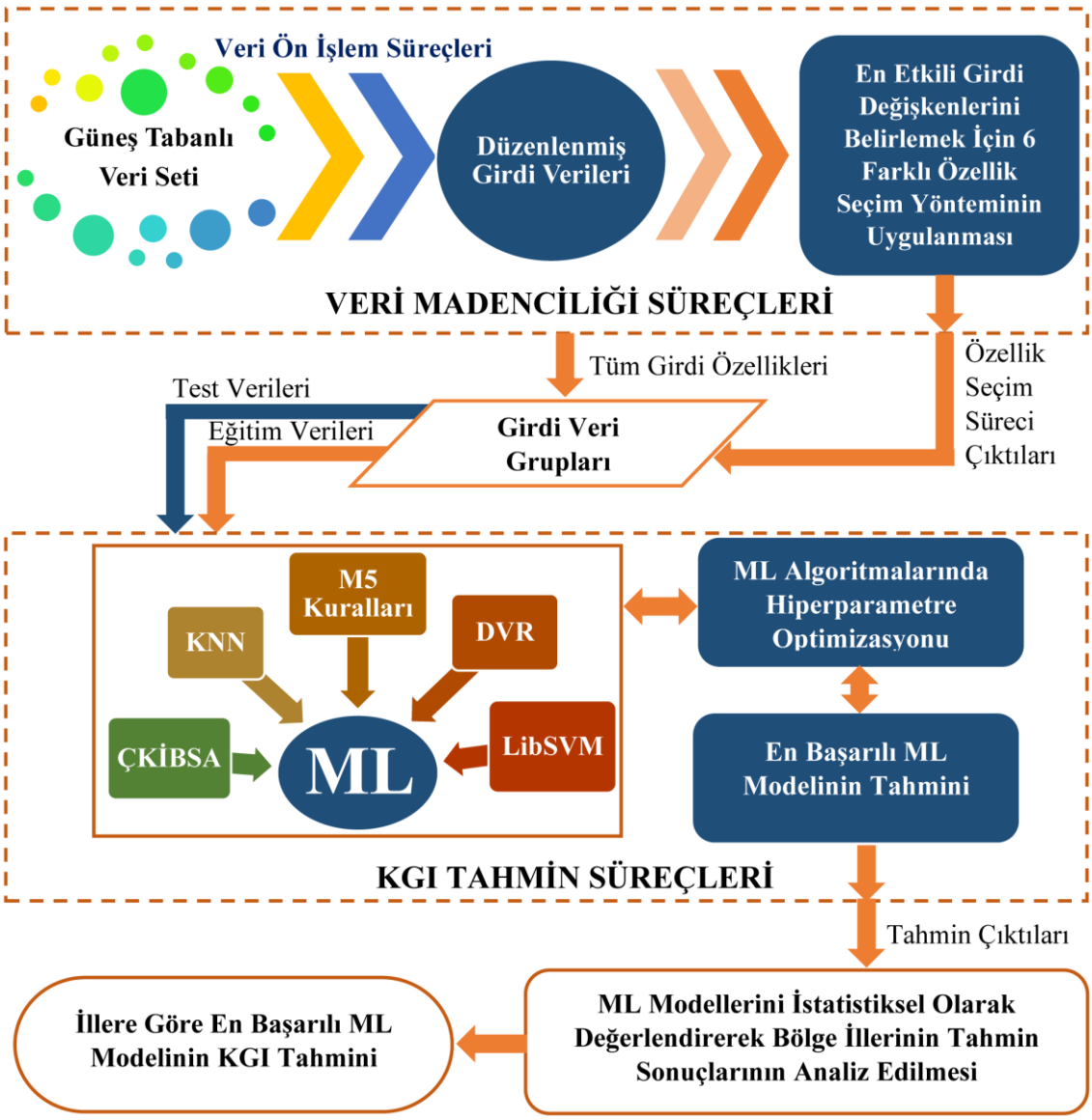
düşen saatlik, günlük ve aylık ortalama KGI verileri kullanılmaktadır. Yapılacak çalışmanın türüne göre bir ön modelleme veya taslak bir tasarım yapılmak isteniyor ise aylık ortalama veriler, kısmen daha kapsamlı bir tasarım yapılacaksa günlük ortalama verilerden yararlanılmaktadır. Ancak, daha hassas ve sonuca odaklı tasarımlarda mutlaka saatlik ortalama veya daha kısa ölçekli verilerin kullanılması gereklidir (Notton ve ark., 2012). Bununla birlikte, saatlik verilerle yapılan çalışmalar, günlük ve aylık verilerle göre tahmin süreçleri daha zor ve karmaşıktır. Bu nedenle saatlik verilerle yapılan güneş ışınım tahmin modellerine daha az rastlanılmaktadır (Zhang ve ark., 2017). Bu tez çalışmasında, KGI tahmin edilmesinde, son yıllarda yaygın kullanımı, kesinliği, güneş tabanlı veri setlerine uyumu ve yüksek performans göstermeleri nedeniyle ML algoritmalarının kullanılması amaçlanmıştır. Ayrıca daha hassas ve sonuca odaklı tahminlere ulaşılmak amaçlandığından saatlik ortalama veriler tercih edilerek veri havuzu oluşturulmuştur.

Günümüzde oldukça fazla alanda uygulaması olan ML algoritmaları veri duyarlı yapılardır. Bu sebepten, ML ile tahmin yapılırken amaca uygun verilerin seçilmesi ve düzenlenmesi çok önemlidir. Makine öğrenmede seçilen girdi verilerinin türü, yapısı, uzunluğu ve birbirleri ile ilişkileri tahmin performansını büyük oranda etkilemektedir. Geliştirilen modellerde temel amaç, kullanılan veri setine göre en başarılı tahmin modelini hesaplamaktır. Veri üzerinde birtakım düzenleme işlemleri yapılsa bile her zaman istenilen ölçekte tahmin kesinliğine ulaşılamayabilir. Değişken sayısı ne kadar fazla olursa tahmin modelinin hedef çıktısı ile ilişkilendirmede problemler yaşanmaktadır. Veri setinde var olan ve girilen her yeni özellik tahmin performansını olumlu veya olumsuz etkileyebilecektir. ML algoritmaları gibi hassas yöntemler ile çalışırken girdinin tüm özelliklerini modellemede kullanmak her zaman arzulandığı gibi üst seviyelerde bir başarı sağlamayabilir. Bu sebeple, veri seti içinden tahmin performansı üzerinde en etkili olanını seçmek gereklidir. Seçim süreci hem tahmin modelin kesinliğine hem de modelin yorumlanmasına büyük katkılar sağlayacaktır ve aynı zamanda modelleme çalışmalarının süresini kısaltacaktır. Güneş ışınımı üzerine yapılan çalışmalarda, KGI tahmin edilmesinde verilerin ön işlem ve seçim prosedürlerinde eksiklikler saptanmış, bu alana çok fazla önem ve zaman ayrılmadığı gözlemlenmiştir. Özellikle son zamanlarda yapılan çalışmalarda meteorolojik, coğrafik ve kategorik değişken sayıları ve veri miktarlarındaki artışlar veriler üzerinde birtakım düzenlemeleri zorunlu kılmaktadır. Veri ön işlemenin önemli adımlarından olan özellik seçim süreçlerinin güneş ışınım tahmininde yaygın kullanılan bir yöntem olmadığı ve

sadece çok az çalışmada farklı yöntem ve tekniklerce uygulandığı gözlemlenmiştir. Bu sebeple veri madenciliği süreçleri yapılan bu çalışmaya dahil edilmiştir. Farklı alanlarda oldukça fazla uygulaması olan ve veri setinden anlamlı, faydalı bilgilerin keşif süreci olarak tanımlanan VM süreçleri sayesinde girdi verilerinin optimizasyonu ve amaca uygun özelliklerin seçilmesi sağlanarak ML algoritmalarına girdi teşkil etmesi hedeflenmektedir. Bu sayede iki farklı yöntem bir uygulamada birleştirilerek kompakt bir şekilde çalışması sağlanmış olacaktır.

Bu tezin amacı, seçilen herhangi bir coğrafik bölge üzerine kurulacak güneş enerjisi santralinin fizibilitesinden önce, KGI değerinin yüksek doğruluk oranlarında, minimum hata ile gerçeğe oldukça yakın bir seviyede tahminini sağlayan modelleri geliştirmektir. Modellerde, tahmin kesinliğini arttırmak, geliştirilen modelin yorumlanmasını kolaylaştırmak ve eğitim süresini kısaltmak amacıyla alanda nadir kullanılan VM süreci destekli ML algoritmalarından faydalanılmıştır. Çalışma alanı olarak Akdeniz bölgesi seçilmiş ve bölgede bulunan sekiz şehrin beş yıllık coğrafik, meteorolojik ve kategorik saatlik verileri kullanılmıştır.

Yapılan çalışma birden fazla adımda incelenmiştir. Çalışmanın ilk aşaması, tahmin modellerinde kullanılacak girdi verilerini çalışmaya uygun hale getirmektir. Bu amaçla veri madenciliği süreçleri içerisindeki temizleme, birleştirme ve dönüştürme adımlarından faydalanılarak, her bir il için yaklaşık 23,442 adet veri çalışmada kullanılır hale getirilmiştir. Çalışmanın ikinci aşaması, güneş ışınımına en fazla etki eden özelliklerin belirlenmesidir. Bu aşamada veri madenciliği süreçlerinde indirgeme adımında kullanılan tekniklerden faydalanılarak filtre ve sarmalayıcı özellik seçim yöntemlerine dayanan altı farklı özellik seçim fonksiyonu ile yaklaşık 29 farklı uygulama geliştirilerek altı farklı girdi veri alt kümesi oluşturulmuştur. Çalışmanın üçüncü aşaması, girdi veri alt kümeleri ve tüm girdi verileri kullanarak ML algoritmaları tabanlı tahmin modelleri geliştirmektir. Bu aşamada asıl amaç literatürde var olan iki farklı yöntem üzerine odaklanmaktır. İlk olarak, yüksek tahmin doğruluğu sebebiyle, YSA tabanlı ÇKİBSA mimarisi kullanılarak Akdeniz bölgesinin tüm şehirleri için tahmin modelleri geliştirilmiştir. İkinci olarak bölge içerisinde seçilen bazı şehirlerin farklı ML algoritmaları ile karşılaştırmalı bir çalışması yapılmıştır. Bu sayede güneş ışınım tahmin çalışmalarında uygulanan yaygın yöntemler, veri-tahmin ve algoritma-tahmin ilişkisi temelinde tek bir çalışmada değerlendirilerek sonuçlar analiz edilmiş ve birbirleri arasında karşılaştırılmıştır. Şekil 1.7'de bu tezde yapılacak çalışmanın metodolojik çerçevesi verilmiştir.



Şekil 1.7 Yapılacak çalışmanın metodolojik çerçevesi (Guher ve ark., 2020)

2. KAYNAK ARAŞTIRMASI

Güneş ışınımının tahmin edilmesi son yıllarda teknolojik gelişmelere paralel olarak üzerinde en fazla çalışma yapılan bir alan olmuştur. Kaynak araştırması, ML ve VM yöntemleri temelli çalışmalar şeklinde iki ayrı başlık altında incelenerek sınıflandırılan araştırmalar paylaşılmıştır.

2.1. ML Yöntemleri Temelli Çalışmalar

ML yöntemleri kullanılarak yapılan yayınlar incelendiğinde, dünyanın farklı coğrafik bölgelerinde yapılacak çalışmanın türüne göre bir, birden fazla veya hibrit ML algoritmalarından faydalanılarak KGI'nın tahmin edilmesi amaçlanmıştır. Yapılan çalışmalarda, veri ve tahmin ilişkisinin iki farklı şekilde uygulandığı sonucuna varılmıştır. Birinci yöntemde meteorolojik ve coğrafik verilerin tamamı kullanılarak ML yöntemi içerisinde güneş ışınımını en iyi tahmin eden algoritmanın veya algoritmaların belirlenmesi hedeflenmektedir. Bu tür çalışmalarda algoritma performansı ön plana çıkarılmıştır. İkinci yöntemde ise ML öğrenme algoritmalarının veri odaklı yapılar olduğundan hareketle, veri ön işleme sürecine önem verilmiş ve veri seti içerisindeki değişkenler farklı şekil ve yöntemlerce seçilip, gruplandırılarak en başarılı ML algoritması modeli bulunmaya çalışılmıştır. Bu sayede güneş ışınımına etki eden en optimal veriler belirlenerek veri-tahmin performansı değerlendirilmiştir. Her iki yöntemde de en fazla kullanılan algoritmanın YSA tabanlı modeller olduğu görülmüştür (Qazi ve ark., 2015).

2.1.1. Algoritma tabanlı modelleme

(Mohandes ve ark., 1998), model sınıflaması, optimizasyon, tahmin ve otomatik kontrol gibi alanlarda yeni yeni uygulaması olan ML algoritmalarından geri-yayılım tabanlı YSA'ndan faydalanarak Sudi Arabistan'ın 41 meteorolojik ölçüm istasyonundan aldıkları enlem, boylam, yükseklik ve güneşlenme süresi girdi verilerini kullanarak aylık ortalama günlük KGI tahmin etmişlerdir. Tahmin edilen değerler ile ölçülen gerçek değerler arasında Ortalama Mutlak Yüzde Hata (MAPE) istatistiksel ölçeğine göre iyi bir uyum olduğu görülmüştür. Benzer şekilde (Al-Alawi ve Al-Hinai, 1998) yaptıkları çalışmada güneş ışınımının doğrudan ölçümünün yapılmadığı yerlerde KGI

tahmin etmek için YSA ML yönteminden faydalandılar. Modelde 8 girdili, 1 çıktılı bir yapı kullandılar. KGI'nın ölçümünün yapıldığı yerlerde toplanan girdi verileri kullanılarak geliştirilen en başarılı modelde %93 tahmin doğruluğuna, ölçümün yapılmadığı bir yere kurulan ve 12 ay ölçüm yapılan girdi verileri aynı sinir ağı modeli ile tahmin yapıldığında %95 tahmin doğruluğuna ulaştılar. Elde edilen sonuçlarla, geliştirilen modelin genelleme ve doğru tahminler üretme yeteneğini gösterdiler.

(Mohandes ve ark., 2000), yaptıkları bir başka çalışmada ML tabanlı Radyal Temelli Fonksiyon Sinir Ağı (RBFNN) YSA mimarisinden faydalanarak aylık ortalama günlük KGI'nı tahmin etmeyi sağlayan modeller geliştirmişlerdir. Ayrıca tahmin sonuçlarını Çok Katmanlı Algılayıcılar (MLP) YSA mimarisi ve klasik regresyon modeli ile karşılaştırmışlardır. Bu çalışmada kullanılan ölçüm verileri Sudi Arabistan'ın 41 meteorolojik istasyonundan alınmakta ve 5 girdili enlem, boylam, yükseklik, güneşlenme süresi ve ay verileri ile tahmin modelleri oluşturulmaktadır. MAPE ölçeğine göre tahmin sonuçlarının, RBFNN ile geliştirilen modellerin MLP ve klasik regresyon modeline göre daha başarılı olduğunu belirttiler.

(Dorvlo ve ark., 2002), yaptıkları çalışmada Umman'ın KGI tahmin etmek amacıyla MLP ve RBFNN ML modellerinden yararlanmışlardır. Hedef bölgede bulunan sekiz farklı ölçüm istasyonundan alınan on yıllık ölçülen 5 girdi verisini kullanarak tahmin için ölçüm istasyonu tabanlı 28 farklı tahmin modeli geliştirdiler. Geliştirilen 28 farklı modelde, 8 ölçüm istasyonundan rastgele seçilen 6 ölçüm istasyonu verisi MLP ve RBFNN modellerine girdi olarak ayarlanırken, 2 istasyonun verileri modelleri test etmek amacıyla kullanılmıştır. Kök Ortalama Kare Hatası (RMSE) istatistiksel ölçeğine göre MLP 24'üncü modelde 1.01, RBFNN ise 28'inci modelde 0.83 doğruluk oranında KGI başarılı bir şekilde tahmin etmiş ve RBFNN modelinin azda olsa MLP'a göre daha iyi performans gösterdiğini hesaplamışlardır.

(Sözen ve ark., 2004), Türkiye'deki güneş enerjisi potansiyelini ML yöntemine dayalı YSA algoritması ile tahmin edilmesine dayalı bir çalışma yaptılar. Geliştirilen modellerde YSA'nın farklı mimari yapılarını kullanmak yerine, Ölçekli Eşlenik Gradyan (SCG), Pola-Ribiere Konjugat Gradyan (CGP) ve Levenberg-Marquardt (LM) gibi 3 farklı öğrenme algoritmasını temel alan tahmin modelleri geliştirdiler. Modelleri eğitmek için Türkiye'nin farklı illerine ait son üç yılda ölçülen meteorolojik ve coğrafik enlem, boylam, yükseklik, ay, ortalama güneşlenme süresi, ortalama sıcaklık ve güneş ışınımı girdi verilerini kullandılar. Belirtme Katsayısı (R^2) ve MAPE ölçeğine göre en başarılı tahmin modeli sırasıyla 0.9989 ve %2.92 ile Antalya ili için bulunmuştur. YSA

algoritmasının, klasik regresyon modellerine göre KGI tahmin etmek için daha uygun bir yöntem olduğunu gösterdiler. Bu tez kapsamında geliştirilen YSA modellerinde farklı öğrenme algoritmalarının performans değerlendirmeleri gerçek tabanlı verilere uygulanarak değerlendirilmiştir. (Sözen ve ark., 2005), bir başka çalışmada Türkiye'nin 12 farklı ilinin bulunduğu girdi verilerini kullanarak YSA tabanlı SCG, LM ve CGP öğrenme algoritmaları destekli tahmin modelleri geliştirdiler. Geliştirilen modellerde 9 ilin girdi verileri eğitim amaçlı, 3 ilin verileri test amaçlı kullanılmış ve 2000-2003 arası ölçülen 4 yıllık meteorolojik ve coğrafik veriler kullanılmıştır. En başarılı KGI tahmin modeli, SCG öğrenme algoritması temelli 6 gizli nöronlu tahmin modelinde Siirt ili girdi verileri kullanılarak R^2 ve MAPE ölçeğine göre sırasıyla 0.9955 ve %6.78 değerleri ile elde edilmiştir. Bu çalışma bir önceki yapılan çalışmada kullanılan YSA'nın güneş ışınımı tahmin etme yeteneğinin doğrulamasını amaçlamışlardır.

(Alam ve ark., 2006), Hindistan'ın farklı iklim koşullarına sahip 11 istasyondan alınan KGI verilerini kullanılarak doğrudan ışınımı değerini tahmin etmeyi amaçladılar. Tahmin modellerinde Çok Katmanlı İleri Beslemeli Geri Yayılım Ağı (MLFFBPN) algoritmasına dayanan bir YSA mimarisi kullandılar. Sonuçlar, 7 ile 11 gizli katman nöron sayılı tahmin modeller içerisinde, 9 gizli kaman nöron sayılı modelin RMSE = 1.65 ile en başarılı doğrudan ışınım tahmin performansına ulaştığını göstermiştir. Bir başka çalışmada, (Mubiru ve Banda, 2008), Uganda'daki 4 farklı il için yatay bir yüzeye düşen aylık ortalama günlük KGI tahmin etmek için MLFFBPN mimarisinde dayanan ML algoritması kullandılar ve deneysel yöntemle geliştirilen matematiksel denklem sonuçları ile karşılaştırdılar. Hedef bölgeden seçilen 3 il verileri MLFFBPN modelini eğitmek, 1 il modeli test etmek amacıyla kullanıldı. Tanjant Sigmoid, Lojistik Sigmoid ve Doğrusal olmak üzere 3 transfer fonksiyonu, SCG, LM, CGP olmak üzere 12 eğitim algoritması ve tek veya iki katmanlı sinir ağı mimarisi kullanılarak geliştirilen modeller eğitildi. Sonuç olarak tek gizli katmanda 15 gizli nöronlu, tanjant sigmoid transfer fonksiyonlu, LM geri yayılım eğitim algoritmalı tahmin modeli, korelasyon katsayısı (R) ve RMSE göre sırasıyla 0.974 ve 0.385 KGI tahmin etmişlerdir. Diğer bir çalışmada (Fadare, 2009), Nijerya'nın 195 şehrinin 10 yıllık aylık ortalama günlük meteorolojik ve coğrafik girdi verilerini kullanarak KGI tahmin etmek amacıyla MLFFBPN mimarisine dayanan tahmin modelleri geliştirmiş ve çalışmanın kolayca uygulanabileceği bir grafik ara yüzü tasarlamıştır. Toplanan 17,550 girdi verisinin 11,700'ü eğitim amaçlı, kalan 5,850'si ise geliştirilen modeli test etmek amacıyla kullanılmıştır. LM öğrenme algoritması kullanılarak geliştirilen, yedi girdili MLFFBPN modeli R = 0.971 değeri ile

KGI'nı başarılı bir şekilde tahmin etmiştir. Bu tez çalışmasında kullanılan tahmin modellerinde YSA'nın MLFFBPN mimarisine hedef alan modeller geliştirilmiştir.

(Wang ve ark., 2011), yaptıkları çalışmada kısa vadeli KGI tahmin edilmesinde geri yayılım algoritmasına dayanan bir sinir ağı mimarisi kullandılar. Sinir ağının yapısı güneş ışınımı tahmin hassasiyetini etkilediği için geliştirilen modellerde karşılaştırma ve çapraz doğrulama yöntemi ile en uygun tahmin modelini seçtiler. 2006 ile 2010 yılları arasında 4 yıllık ölçülen saatlik güneş ışınım verilerini kullanarak iki farklı sinir ağı modeli geliştirdiler. İlkinde 06:00-08:00 saatleri arası 15 girdi verisi, ikincisinde 24 saat ölçüm aralığına dayanan 24 girdi verisi kullandılar. Çift gizli katmanlı ve ilk katmanda 18, ikinci katmanda 13 nöronlu sinir ağı modeli R^2 = 0.9912 ile kısa dönem güneş ışınımını başarılı bir biçimde tahmin etmişlerdir. (Hasni ve ark., 2012), yaptıkları çalışmada saatlik olarak ölçülen meteorolojik verilerini kullanarak Cezayir'in sıcaklık, bağıl nem, ay, gün ve saat girdi verileri ile KGI tahmin etmek amacıyla ML algoritmalarından YSA destekli modeller geliştirdiler. Saatlik olarak ölçülen 3,475 verinin 2,824 tanesi yaklaşık %81 eğitimde, kalan 651 tanesi yaklaşık %19'u modeli test etmek amacıyla ayarlandı. R^2, RMSE. Ortalama Mutlak Hata (MAE) istatistiksel ölçeklere göre ölçülen ve tahmin edilen değerler arasında sırasıyla 0.9999, 2.9971, 0.1721 değerleri ile gerçek ölçülen verilere oldukça yakın bir benzeşim olduğunu sonucuna vardılar.

(Khatib ve ark., 2012), Genelleştirilmiş Regresyon Sinir Ağı (GRNN), Basamak Korelasyon Sinir Ağı (CFNN), MLFFBPN, Elman Geri Yayılım Sinir Ağı (ELMNN) olmak üzere yaptıkları çalışmada dört farklı YSA mimarisini kullanarak saatlik KGI ve dağıtılmış ışınımın tahmin edilmesini amaçlarken, aynı zamanda farklı sinir ağı mimarileri arasında karşılaştırmalı bir çalışma yapmışlardır. Yapılan çalışmada tüm sinir ağlarında farklı aylara ait ölçülen enlem, boylam, sıcaklık, güneşlenme süresi, nem, ölçüm yapılan ay, ölçümün yapıldığı gün ve saat girdi verilerini kullanarak 8 girdili, 2 çıktılı modeller geliştirdiler. MAPE, RMSE ve Ortalama Sapma Hatası (MBE) hata istatistiksel ölçeklerine göre GRNN ile geliştirilen model sırasıyla %4.29 5.75 ve 0.29 değerleri ile diğer YSA mimarilerinden daha iyi performans gösterdiğini hesaplamışlardır.

(Zeng ve Qiao, 2013) yaptıkları çalışmada, Amerika Birleşik Devletleri'nin Seattle, Denver ve Miami bölgelerinin kısa vadeli güneş enerjisi tahmini için en küçük kare DVM tabanlı ML modeli geliştirdiler. Geliştirilen modelleri AR ve RBFNN tabanlı

modellerle karşılaştırdıklarında DVM'nın daha iyi bir performans elde ettiğini ifade ettiler.

(Dahmani ve ark., 2014), bir saatten kısa aralıklarla ölçülen küçük zaman adımlı ve ortalaması alınmamış meteorolojik veriler kullanılarak KGI tahmin edilmesine odaklandılar. Deneysel ve istatistiksel yöntemlerin kısa zaman aralıklarında ölçülen değerler için etkili olmadığından hareketle bu alanda oldukça nadir kullanılan YSA algoritmasından faydalandılar. Yaptıkları çalışmada, Cezayir'in Bouzareah şehrindeki ölçüm istasyonundan 5 dakikalık aralıklarla 2 yıl boyunca ölçülen düzenlenmiş 75,674 adet veri ile modeller geliştirdiler. Girdi olarak deklinasyon açısı, zenit açısı, azimut açısı, dünya dışı radyasyon ve 5 dakikada bir ölçülen KGI, çıktı olarak 36.8 derecelik eğimli bir yüzeye düşen KGI ışınımı verilerini kullanarak normalleştirilmiş RMSE = %8.81, R^2 = 0.994, MAE = 2.41 ile iyi bir doğruluk oranı tahmin etmişlerdir. (Kadirgama ve ark., 2014), yaptıkları çalışmada Malezya'nın güneş ışınım potansiyelini tahmin etmek için hızlı yayılma algoritmasına dayanan YSA modelleri geliştirmişlerdir. 10 dakikalık kısa dönem ölçüm zaman aralıklarıyla toplanan 1 aylık 1,617 veri ile 8 girdili, 20 gizli katman nöronlu geliştirilen sinir ağı modeli güneş ışınımını MAPE = %7.74 ve R^2 = 98.9 değerleriyle tahmin etmişlerdir.

(Salcedo-Sanz ve ark., 2014), güney İspanya'nın Murcia kentinde bulunan farklı ölçüm istasyonlarından alınan iki yıllık meteorolojik verileri kullanarak ML algoritmalarından Aşırı Öğrenme Makinesi (EML) yöntemlerinden faydalanarak günlük KGI tahmin etmeyi amaçladılar. Mercan Resifleri Optimizasyonu ile yapılan hibrit çalışmada, EML'nın geliştirilen modelin tahmininde elde edilen çözümleri iyileştirmek adına sinir ağının ağırlıklarının geliştirilmesini hedeflediler. Geliştirilen hibrit modelin klasik EML ve DVR ile karşılaştırıldığında daha iyi performans gösterdiğini belirttiler.

(Lazzaroni ve ark., 2015), İtalya'nın Milano şehrinde üç yıllık saatlik verileri kullanarak Otoregresif modeller, DVR ve EML ML algoritmaları ile saatlik, günlük ve aylık zaman dilimlerine göre geliştirilen KGI tahmin modellerini K-NN algoritması ile karşılaştırdılar. Yaptıkları çalışmada, zaman serilerini modellemek için 3 yıllık saatlik ölçülen girdi veri seti kullanılarak geliştirilen en başarılı modelin EML ve DVR'na göre MAPE ölçeğinin %5'in altında olduğunu gösterdiler.

(Karabacak, 2016) yaptığı çalışmada, rüzgâr-fotovoltaik hibrit güç sistemlerini tanılama ve kontrol amaçlı bir YSA modeli tasarlanmış, tasarlanan sinir ağı modeli bir mikrobilgisayara gömülerek örnek bir hibrit güç sistemini kontrol amaçlı kullanmıştır. Yapılan çalışma ile YSA tabanlı kontrol algoritması özgün olarak modellenmiştir.

(Premalatha ve Valan Arasu, 2016), yaptıkları çalışmada Hindistan'ın farklı iklim koşullarına sahip beş farklı konumun aylık ortalama KGI tahmini için her biri dört farklı eğitim algoritmasına sahip iki farklı YSA modeli kullandılar. İlk YSA modelinde 5 farklı konumun 4'nün verileri eğitim, 1'i test için kullanılırken, ikinci YSA modelinde eldeki tüm girdi verilerinin %80'ni eğitim, %20'si test olarak ayarlanmıştır. Geliştirilen modellerde 9 girdi verisi ve 4 farklı geri yayılım algoritması ile KGI değeri R = 0.9545, MAE = 0.7800 ile LM öğrenme algoritmasının kullanıldığı ikinci YSA modeli için tahmin edildiğini hesaplamışladır.

(Wang ve ark., 2016) yaptıkları çalışmada, 1961-2014 tarihleri arasında, hava sıcaklığı, hava basıncı, bağıl nem, su buharı basıncı, saatlik güneşlenme süresi gibi günlük gözlenen meteorolojik verileri 12 farklı iklim özelliğine sahip istasyondan almışlardır. Günlük KGI'nı etme amacıyla MLP, GRNN ve RBFNN sinir ağı mimarisine dayanan 3 farklı YSA modelleri için karşılaştırmalı bir çalışma yaptılar. MLP ve RBFNN tabanlı modellerin KGI tahmin etmede GRNN modeline göre doğruluk oranlarının daha kesin olduğunu gösterdiler. (Wang ve ark., 2017) Çin'in 21 meteorolojik ölçüm istasyonundan aldıkları günlük meteorolojik değişkenleri kullanarak yaptıkları diğer bir çalışmada, iki farklı optimize edilmiş Uyarlanabilir Nöro-Bulanık Çıkarım Sistemi (ANFIS) YZ yöntemi ve M5 Karar Ağaçları ML algoritmasını kullanarak günlük KGI tahmin etmeyi amaçladılar. Tahmin sonuçlarını deneysel yöntem kullanılarak geliştirdikleri bir başka model ile karşılaştırdıklarında ANFIS tabanlı iki modelin hem M5 hem de deneysel yöntemle geliştirilen modelden daha iyi bir tahmin performansına ulaştığı sonucuna varmışlardır.

(Hassan ve ark., 2017) yaptıkları çalışmada, Doğrusal Regresyon ve DVR ML algoritmaları ile Fiji'deki Nadi Havaalanının bulunduğu lokasyonda 2015 yılına ait girdi verilerini kullanarak güneş ışınımını tahmin ettiler. DVR algoritmasında RBF çekirdeğini ve çekirdek parametresinde sayısal optimizasyon yaparak 3 farklı model geliştirdiler. Doğrusal Regresyon ve DVR modellerinde geliştirilen en başarılı modelin MAE istatistiksel ölçeğine göre değerini sırasıyla 2.203 ve 0.580 olarak bulmuşlardır.

(Hamdan ve ark., 2017) bu çalışmada Ürdün'ün Amman coğrafik bölgesinde saatlik KGI'nı tahmin etmek için MLFFBPN, ELMNN ve Doğrusal Olmayan Dışsal Girdili Otoregresif Ağ (NARX) gibi YSA tabanlı üç farklı mimariyi kullandılar. NARX modelinin ELMNN ve MLFFBPN ağlarına kıyasla tahmin sürecinde en başarılı model olduğunu $R^2 = 0.9892$ değeri ile göstermişlerdir.

(Bou-Rabee ve ark., 2017), Kuveyt'in 5 farklı bölgesinden alınan girdi verilerini kullanarak günlük ortalama KGI'nı tahmin etmek amacıyla bir model geliştirdiler. Tahmin modelini geliştirmek için YSA ML algoritmasından faydalanarak farklı girdi ve gizli katman nöron sayılarının optimize edilmesi ile 3 farklı modelin performanslarını karşılaştırdılar. Geliştirilen en başarılı tahmin modeli %94.75 oranında Kuveyt'in güneş ışınımını tahmin ettiğini ifade etmişlerdir.

(Khosravi ve ark., 2018a) yaptıkları çalışmada, güney İran'ın Abu Musa Adası'nın saatlik KGI'nı tahmin etmek amacıyla, Bulanık Çıkarım Sistemi (FIS), ANFIS, RBFNN, MLFFBPN ve DVR ML algoritmalarını kullanarak modeller geliştirilmiş ve birbiri ile karşılaştırmışlardır. Tahmin modellerinde iki tür girdi veri grubu kullanmışlardır. İlk girdi veri grubunu, zaman, sıcaklık, basınç, rüzgâr hızı ve bağıl nem ile oluştururken, ikinci veri grubu gelecekteki değerleri tahmin etmek için geçmiş değerlerin kullanıldığı zaman serisi verileri ile oluşturmuşlardır. İlk girdi verileri kullanılarak en iyi sonuç DVR ML algoritmasıyla R = 0.9999 ile alınırken, bunu MLFFBPN algoritmasıyla geliştirilen modelin R = 0.9795 ile takip ettiğini gösterdiler. İkinci girdi verileri kullanılarak en iyi sonuçlar DVR, MLFFBPN ve ANFIS makine öğrenme algoritmasıyla birbirine oldukça yakın R = 0.95 düzeyinde elde edildiğini hesaplamışlardır.

(Yıldırım ve ark., 2018) yaptıkları çalışmada, Türkiye'nin Doğu Akdeniz Bölgesi günlük ortalama KGI tahmin etmek amacıyla YSA ve Angström-Prescott modellerinin performanslarını karşılaştırmayı amaçladılar. Modelleme çalışmasında kullanılan 11 yıllık verileri, 1997'den 2007 yılları arasında ölçülen Adana, Mersin, Hatay ve Kahramanmaraş illerine ait ölçüm istasyonlarından sağladılar. Seçilen verilerin 9 yılını eğitim amaçlı, 2 yılını modeli test etmek amacıyla düzenlediler. İstatistiksel test ve hata analiz sonuçlarına göre YSA tarafından geliştirilen modelin %96 doğrulukta güneş ışınımını tahmin ettiğini, deneysel yöntemlerin tahmin performansının oldukça düşük kaldığını ifade etmişlerdir.

(Ozoegwu, 2019) bu çalışmada, NAR, NARX ve hibrit (YSA-NAR) tabanlı üç farklı ML algoritması kullanarak KGI'nın aylık ortalama günlük değerlerini tahmin etmeyi amaçladılar. Nijerya'nın 7 coğrafik bölgesinin verileri kullanılarak yapılan çalışmada, hibrit YSA tabanlı model diğer iki yönteme göre geliştirilen tahmin modellerine göre R^2 = 0.96 ve RMSE = 0.33 ile oldukça iyi bir performans göstermişlerdir.

(Feng ve ark., 2019) yaptıkları çalışmada, hava sıcaklığını girdi verisi olarak kullanıp KGI tahmini için farklı ML ve deneysel modelleri karşılaştırdılar. Bu amaçla, dört ML algoritması YSA, Rastgele Orman (RF), Dalgacık Dönüşümü Sinir Ağı (DNN) ve hibrit Evrimsel Algoritma-YSA ve dört sıcaklık tabanlı deneysel model (Hargreaves-Samani, Bristow-Campbell, Jahani ve Fan) Çin'in ılıman bölgelerinde günlük KGI tahmin etmek için uyguladılar. Tahmin sonuçlarına göre, hibrit modelin diğer ML ve deneysel modellerle karşılaştırıldığında YSA ile birlikte daha iyi tahmin kesinliğine ulaştığını hesaplamışlardır.

(Şahin, 2019) yatığı çalışmada, Sivas ilinin aylık ortalama toplam KGI'nı tahmin etmek için 5 farklı ölçüm istasyonundan elde edilen veriler kullanılarak YSA ML algoritması ile modeller geliştirmiştir. YSA yöntemi kullanılarak tahmin edilen güneş ışınımı tahmin sonuçlarının toplanan ölçüm verileriyle iyi bir uyum içerisinde olduğunu gözlemlemiştir.

(Alsafadi, 2020), Eskişehir bölgesinde saatlik KGI'nı tahmin etmek için ML yöntemlerinin performanslarını değerlendirmek amacıyla beş deneysel model ile karşılaştırmasını yapmıştır. YSA, DVR ve Regresyon Ağaçları (RA) tabanlı geliştirilen tahmin modellerinden, DVR modeli R = 0.99 ile hem diğer ML algoritmalarından hem de deneysel yöntemlerden daha yüksek tahmin doğruluğuna ulaştığı sonucuna varmışlardır.

(Pang ve ark., 2020) KGI tahmini için yaptıkları çalışmada, YSA ve Tekrarlayan Sinir Ağı (RNN) tabanlı Uzun Kısa Süreli Bellek (LSTM) derin öğrenme algoritmasın dayanan modeller geliştirmiş ve performanslarını karşılaştırmışlardır. Alabama'daki yerel bir hava istasyonundan alınan 10, 30, 60 dakikalık zaman dilimlerinde ölçülen meteorolojik verileri modellerin eğitim ve test süreçlerinde kullandılar. 10 dakikalık zaman diliminde ölçülen girdi verisi kullanılarak YSA ve RNN algoritmaları ile geliştirilen tahmin modelleri için R^2 ölçeğine göre sırasıyla 0.974 ve 0.983, 30 dakikalık girdi verisi ile 0.97 ve 0.977 ve 60 dakikalık girdi verisi ile 0.933 ve 0.97 tahmin sonuçlarına ulaşmışlardır.

2.1.2. Veri tabanlı modelleme

(Benghanem ve ark., 2009) yaptıkları çalışmada, Suudi Arabistan'ın Medine şehrindeki günlük ortalama KGI tahmin modellerinde 1998-2002 yılları arası sıcaklık, güneşlenme süresi, bağıl nem, yılın günü verileri girdi ve güneş ışınımı çıktı olarak

kullanılarak girdi değişkenleri temelli 6 farklı YSA modeli geliştirerek hem deneysel yöntemlerle hem de kendi aralarında karşılaştırdılar. Her bir veri değişkeninin tek başına ve farklı kombinasyonları denenerek en başarılı tahmin modelini bulmaya çalıştılar. Güneşlenme süresi ve hava sıcaklığı kullanılarak oluşturulan modelin R = 0.9765 ile diğer modellere nazaran daha başarılı bir performans gösterdiğini belirttiler.

(Moghaddamnia ve ark., 2009), ANFIS, Local Linear Regression (LLR), Doğrusal Olmayan Dışsal Girdili Otoregresif Ağ (NARX), MLP ve ELMNN ML algoritmalarını kullanarak İngiltere'nin Brue havzasının KGI'nı tüm girdi verilerinden seçilen farklı meteorolojik girdi değişkenleri ile tahmin etmeyi amaçladılar. Girdi değişkenlerinin seçiminde gama testini kullanarak alternatif bir yaklaşım sundular ve böylece deneme yanılma iş yükünü büyük ölçüde azaltmayı hedeflediler. Seçilen girdi değişkenleri yardımıyla geliştirilen 5 farklı modelden LLR algoritması ile geliştirilen modelin performansı RMSE = 22.4 ve R^2 = 0.90, MLP algoritması ile geliştirilen modelin performansı RMSE = 28.1 ve R^2 = 0.83 ile diğer modellere göre daha üstün performansa ulaştığını ifade etmişlerdir

(Paoli ve ark., 2010) yaptıkları çalışmada, optimize edilmiş MLP, ARIMA, K-NN, Markov Zincirleri ve Bayesci Çıkarım gibi ML algoritmalarından faydalanarak Fransa'nın Korsika Adasının 19 yıllık ölçülen verileri ile KGI tahmin ettiler. Verilerin değişkenlerin seçimi ve optimizasyonunda Bayesci Çıkarım ve Markov zincirlerinden faydalandılar. MLP algoritması ile geliştirilen tahmin modelinin RMSE = 0.385 ile daha iyi bir tahmin kesinliğine ulaştığını göstermişlerdir.

(Behrang ve ark., 2010), 2002-2006 arası İran'nın Dezful şehrinin meteorolojik 6 farklı girdi verisinin 6 farklı kombinasyonlarına dayanan, MLP ve RBFNN tabanlı YSA modeller geliştirerek günlük KGI'nı tahmin etmişlerdir. En başarılı tahmin modelleri MLP sinir ağı için MAPE %5.21 olarak hesaplanırken, RBFNN için %5.56 olarak bulunmuştur.

(Al Shamisi ve ark., 2011) Birleşik Arap Emirlikleri'nin Al Ain şehrinin günlük ortalama KGI tahmin edebilmek amacıyla sıcaklık, ortalama rüzgâr hızı, güneşlenme süresi, ortalama bağıl nem meteorolojik verilerinden yararlanarak deneme yanılma yolu ile 4 girdili 1 tane, 3 girdili 4 tane, 2 girdili 6 tane toplamda 11 değişken kombinasyonuna dayanan MLP ve RBFNN mimarisi temelli YSA tahmin modelleri geliştirdiler. 1995-2007 arasında günlük ölçülen verilerden, 1995-2004 arası ölçülenler ağların eğitiminde, 2005-2007 arası ölçülenler ise modelleri test edebilmek amacıyla kullandılar. Her iki YSA algoritmasında 14 farklı geri-yayılım öğrenme algoritması

kullandılar. En iyi tahmin sonuçlarını MLP modelleri için R^2 ve RMSE göre sırsıyla 0.95 ve 0.351 ile sıcaklık, güneşlenme süresi ve bağıl nem özellikleri gruplanarak bulunurken, RBFNN modelleri için benzer şekilde aynı veri özellik grubuyla 0.93 ve 0.446 değerlerini hesaplamışlardır.

(Chen ve ark., 2011) yaptıkları çalışmada, hava sıcaklıklarını kullanarak güneş ışınım tahmininde nispeten yeni bir ML algoritması olan DVM'ni kullanarak aylık ortalama günlük KGI tahmin etmeyi amaçladılar. 1973'ten 2000 yılına kadar ölçülen 28 yıllık veriden maksimum sıcaklık (Tmak), minimum sıcaklık (Tmin) ve maksimum-minimum sıcaklık (Tmak-Tmin) arasındaki farkı kullanarak elde etikleri 3 değişkenden 7 farklı girdi kombinasyonu oluşturarak doğrusal, polinom ve radyal tabanlı çekirdek fonksiyonuna göre 21 tahmin modeli geliştirdiler. Polinom çekirdek fonksiyonun kullanıldığı (Tmak, Tmin) girdi verileri grubu ile en başarılı DVM modeli RMSE = 0.8330 ile KGI'nı başarılı olarak tahmin etmişlerdir.

(Lazzús ve ark., 2011), Şili'nin La Serena şehrindeki 2001 ile 2007 yılları arasında toplanan saatlik KGI'nın tahminini yapmak amacıyla YSA algoritmasına dayanan modeller geliştirdiler. Girdi veri özelliklerinin güneş ışınımına etkisini hesapladıktan sonra farklı değişken kombinasyonlarına göre modeller geliştirdiler. En başarılı modelin girdi veri özellikleri rüzgâr hızı, bağıl nem, hava sıcaklığı ve toprak sıcaklığından oluştuğunu ve KGI'nı $R^2 = 0.9437$ ile tahmin etmişlerdir.

(Ozgoren ve ark., 2012), Türkiye'nin 31 farklı ilindeki meteoroloji istasyonundan 2000-2006 yılları arasında elde edilen 2074 adet aylık ortalama günlük verileri kullanılarak KGI'nın tahmin edilmesini amaçladılar. Çalışmada Çoklu Doğrusal Olmayan Regresyon (MNLR) ve YSA ML algoritmalarından faydalanarak modeller geliştirip ve karşılaştırdılar. Aynı zamanda 15 girdi verisi içerisinde en optimal girdi verilerini seçmek amacıyla "Enter" regresyon yöntemi kullanılarak 10 girdi verisi belirlediler. Stepwise yöntemi ile 10 farklı MNLR modeli geliştirip ve karşılaştırma amacıyla yine aynı girdi verileri ile 10 farklı ANN modeli oluşturdular. En iyi performans R = 0.9936 ile ANN-10 modelinden elde edildiğini ifade etmişlerdir.

(Benmouiza ve Cheknane, 2013), denetimsiz K-Ortalamalar kümeleme algoritması ve YSA'nın kombinasyonuna dayalı olarak saatlik KGI'nın tahminini amaçladılar. K-Ortalamalar ile girdi veri uzayını kümelediler, ardından NAR ağı ile her bir kümeyi modelleyerek KGI'nı tahmin edip, ARMA yöntemi ile karşılaştırdılar.

(Long ve ark., 2014) yaptıkları çalışmada, ML tabanlı YSA, K-NN, DVM ve Çok Değişkenli Doğrusal Regresyon (MLR) algoritmalarını kullanılarak günlük

ortalama KGI'nının tahmini araştırdılar ve modellerin karşılaştırmalı bir analizini yaptılar. Girdi veri setinden en önemli parametrelerin seçiminde doğrusal regresyon, hız regresyonu ve DVR algoritmalarını kullanarak 0,5'e eşit ve üstündekileri belirleyip her üç yöntem sonunda bu şarta uyanları seçtiler. Belirlenen girdi verilerinin YSA, K-NN, DVM ve MLR algoritmasının çalışma yapısına uygun olacak şekilde seçiminde bir algoritma geliştirerek her algoritma için farklı veri değişkenlerini belirlediler. Ardından 4 farklı algoritma için tahmin modelleri geliştirip karşılaştırdılar. YSA modeli diğer algoritmalara göre yaklaşık %90 ile KGI başarılı bir biçimde tahmin ettiğini gösterdiler.

(Hejase ve ark., 2014), Birleşik Arap Emirlikleri'nin üç büyük şehri olan Abu Dabi, Dubai ve Al-Ain için KGI'nı tahmin etmek amacıyla YSA tabanlı MLP ve RBFNN mimarilerini dayanan tahmin modellerini kullandılar. Meteorolojik ölçüm istasyonlarından topladıkları 4 girdi verisini farklı şekilde seçip gruplayarak 11 farklı tahmin modeli geliştirdiler. Abu Dabi şehrinde sıcaklık, bağıl nem ve güneşlenme süresi veri grubu ile geliştirilen RBFNN tabanlı tahmin modeli R^2 = 0.94 ve MAPE = %4.06 ile, Al-Ain şehrinde tüm girdi verileri kullanılarak geliştirilen MLP tabanlı tahmin modeli R^2 = 0.92 ve MAPE = %4.61 ile, Dubai şehrinde sıcaklık, rüzgar hızı ve güneşlenme süresi veri grubu ile geliştirilen RBFNN tabanlı tahmin modeli R^2 = 0.95 ve MAPE = %4.09 ile en iyi tahmin modeli olarak hesapladılar.

(Olatomiwa ve ark., 2015) yaptıkları çalışmada, ML tabanlı DVM algoritmasını kullanarak KGI'nı tahmin etmek amacıyla bir model geliştirdiler. Modelde, Nijerya'nın üç farklı bölgesinden alınan 21 yıllık aylık ortalama günlük güneşlenme süresi, minimum sıcaklık ve maksimim sıcaklık meteorolojik verilerini kullandılar. Tahmin performansını arttırmak amacıyla girdi verilerinden en uygun özelliklerin seçiminde Ateşböceği Algoritmasından (FFA) faydalandılar. DVM-FFA modeli R^2 = 0.802 ile YSA ve Genetik Algoritma (GA) tahmin sonuçları ile karşılaştırıldığında en başarılı tahmini gösterdiğini belirttiler.

(Ahmad ve ark., 2015), Yeni Zelanda'nın en büyük şehri Auckland'ın saatlik ölçülen 8 farklı meteorolojik girdi verileri kullanılarak KGI 24 saat önceden tahmin etme yeteneğini araştırmayı amaçladılar. Bu amaçla NARX, MLP, ARMA ve kalıcılık ML algoritmalarına dayanan modeller geliştirdiler. KGI tahmininde en önemli uygulamalardan biri girdi değişkenlerinin seçimi olmasından ve değişkenlerin seçiminin veri kalitesine ve korelasyonuna bağlı olmasından hareketle tüm girdi verilerinin güneş ışınımına etki korelasyonunu hesapladılar. Önem seviyelerine göre toplamda 12 farklı girdi veri grubu oluşturdular. NARX mimarisine göre 7 girdili geliştirilen YSA

algoritması diğer ML modellerinden RMSE = 0.243 değeri ile daha üstün performans gösterdiğini belirttiler.

(Piri ve ark., 2015), İran'ın farklı iklim koşullarına sahip Zahedan ve Bojnurd bölgelerinde güneş ışınımını tahmin etmek amacıyla polinom ve radyal temelli çekirdek fonksiyonları kullanarak DVR algoritmasına dayanan ML tabanlı modeller geliştirdiler ve deneysel yöntemlerle karşılaştırdılar. Minimum sıcaklık, maksimum sıcaklık, bağıl nem ve güneşlenme süresi girdi verilerini geliştirilen modellere deneme yanılma yolu ile gruplandırılarak düzenlediler. Bojnurd bölgesinde 4 girdili polinom çekirdek fonksiyonlu geliştirilen en iyi model için R^2 ve RMSE istatistiksel ölçeklerine göre sırasıyla 0.83 ve 38.85 değerleri hesaplanırken, Zahedan bölgesi için 3 girdili radyal temelli çekirdek fonksiyonlu model için sırasıyla 0.72 ve 49.90 değerlerini hesapladılar. Modellerin karşılaştırmasında DVR modellerinin KGI'nı tahmin etmede deneysel modellerden daha iyi performans gösterdiğini belirttiler.

(Yıldırım, 2016), Doğu Akdeniz Bölgesinde KGI'nı seçilen meteorolojik değişkenlere bağlı olarak tahmin edilmesi amacıyla farklı ölçüm zaman aralıklarına dayanan modeller geliştirmiştir. Girdi verileri ayrı ayrı veya hepsi bir arada kullanılarak en kesin modeller oluşturulmaya çalışılmıştır. Deneysel yöntemler ve YSA algoritmaları kullanılarak geliştirilen modeller kendi aralarında karşılaştırılarak KGI başarılı bir şekilde tahmin edilmiştir.

(Zou ve ark., 2016) yaptıkları çalışmada, Güneydoğu Çin'deki KGI'nı tahmin etmek için uzamsal enterpolasyona dayanan YSA algoritması geliştirildiler. Deneysel yöntemlerden geliştirilen Bristow – Campbell ve Ångström – Prescott tabanlı modelleri, KGI thmininda en iyi modeli bulmak amacıyla YSA modelleri ile karşılaştırdılar. YSA'nda model girdilerinin KGI ile korelasyonlarına göre 4 farklı meteorolojik değişken kombinasyonları belirlenmiş ve 9 girdili 4. model R^2 = 0.91 ve RMSE = 1.34 ile en iyi performans sonucuna ulaşmıştır. Geliştirilen iki deneysel model ile karşılaştırıldığında, 4. model 10 istasyonda KGI'nı tahmin etmede daha yüksek performans gösterdiğini ifade ettiler.

(Azimi ve ark., 2016) yaptıkları çalışmada, güneş ışınım potansiyelini en iyi şekilde tahmin edebilmek ve kesinliğini arttırabilmek için TB K-ortalamaları adında, K-ortalamaları kümeleme algoritması tabanlı YSA destekli hibrit bir yapı kullandılar. K-ortalamaları tabanlı kümeleme algoritmaları, her çalıştırmada farklı ağırlık merkezlerinden küme seçimi yüzünden kesin bir yanıt sağlama yeteneğine sahip olmadığını belirttiler. Bu sebepten önerilen kümeleme algoritması her çalıştırmada sabit

ve kesin sonuçları sağladığını ifade ettiler. Önerilen kümeleme yöntemi ile meteorolojik verileri, girdi kümeleri halinde gruplamak için kullandılar ve bu grupları daha sonra YSA modellerine girdi olarak ayarladılar. Yeni geliştirilen sınıflama algoritması ve MLFFBPN ortaklaşa bir şekilde kullanılması ile farklı zaman dilimleri için güneş radyasyonunu tahmin edilmesinde başarılı bir model geliştirdiler. Önerilen hibrit modelin diğer modellerden daha üstün bir performans sergilediği sonucuna ulaştılar.

(Chiteka ve Enweremadu, 2016), ML tabanlı YSA algoritmasını kullanarak Zimbabwe'nin KGI potansiyelini tahmin ettiler. Bu amaçla 2004-2005 yılları arasında iki yıllık bağıl nem, basınç, açıklık indeksi, ortalama sıcaklık, rakım, enlem ve boylam gibi yedi farklı coğrafik ve meteorolojik verilerden faydalandılar. Bu verileri deneme yanılma yolu ile çeşitli şekilde gruplandırılarak YSA ile modeller geliştirmiş ve en iyi performansı $R^2 = 0.99$ ile bulmuşlardır. Veriler içinde sonuca en fazla etki eden verinin açıklık indeksi ve sıcaklık verileri olduğunu belirttiler. Bir başka çalışmada (Çelik ve ark., 2016), Türkiye'nin Doğu Akdeniz Bölgesi'nde bulunan Adana, Mersin, Antakya, Osmaniye ve Kahramanmaraş illerinin etkili bir aylık ortalama günlük KGI tahmini gerçekleştirmek amacıyla optimize edilmiş YSA modelleri geliştirdiler. Tahmin modellerinin performanslarını iyileştirmek amacıyla, giriş özelliklerinin sayısı ve kalitesi, öğrenme algoritmaları, gizli nöron sayıları dikkate alınarak farklı sinir ağı modelleri oluşturdular. Deneme yanılma yolu ile en iyi girdi kombinasyonunun seçildiği 4 ve 6 girdili iki modeli, iki farklı öğrenme algoritması temelinde geliştirdiler. 4 girdili LM öğrenme algoritması ve gizli katmanda 20 nöronlu YSA modelinin elde edilen istatistiksel test sonuçlarına göre, R^2 değeri 0.99'dan fazla ve MAPE değerinin yaklaşık %5 civarında hesaplamışlardır.

(Belaid ve Mellit, 2016), Cezayir'in Ghardaïa şehrinde günlük ve aylık ortalama KGI potansiyelini tahmin edebilmek amacıyla DVM algoritmasından faydalandılar. 6 girdi verisinin farklı kombinasyonlarına göre 42 farklı DVM tabanlı tahmin modelleri geliştirdiler. Günlük tahminde 28 numaralı model 4 girdi verisi ile MAPE = %10.43 hesaplanırken, aylık tahminde 30 numaralı model 2 girdi verisi ile MAPE = %8.40 olarak hesaplanmıştır. Geliştirilen modelleri YSA tabanlı MLP algoritması ile karşılaştırdıklarında DVM'nin daha üstün bir performansa ulaştığını hesaplamışlardır.

(Alsina ve ark., 2016), YSA ML öğrenme algoritmasını kullanarak İtalya'nın Aylık Ortalama Günlük KGI'nın tahminini hedeflediler. 45 farklı coğrafik yerden gelen 13 girdi parametresinden Otomatik İlişki Belirleme (ARD) yöntemi kullanarak seçilen en alakalı değişkenler ile tahmin modellerinin eğitim ve test veri setlerini oluşturdular.

Tüm girdi verileri kullanılarak geliştirilen en iyi modelin MAPE ölçeğine göre performansı %2.66 iken, seçilen 7 girdi verisiyle geliştirilen en iyi modelin performansı %1.67 olarak iyileştiğini belirtmişlerdir.

(Loutfi ve ark., 2017b) yaptıkları çalışmada, MLP ve NARX algoritmaları ile Fas'ın Fes şehrinde 2010-2014 arası beş yıllık ölçülen 9 farklı girdi değişkeni yardımıyla deneme-yanılma yöntemi ile 10 farklı saatlik ortalama KGI tahmin modelleri geliştirdiler. Geliştirilen modeller arasından en başarılı tahmin modelinin 5 girdili NARX ile geliştirmişlerdir.

(Almaraashi, 2018), çok sayıda girdi parametresinin tahmin süreçlerini etkilediğinden hareketle, toplanan veri setindeki değişkenlerin etkili ve verimli bir özellik seçim yöntemi ile seçilmesi üzerine odaklandı. Bu amaçla, YSA algoritmasını kullanarak Suudi Arabistan'ın farklı yerlerindeki günlük KGI'nı tahmin etmek amacıyla Relief algoritması, Monte Carlo algoritması, rastgele kurbağa algoritması ve Laplacian skor algoritması gibi 4 farklı özellik seçme yöntemlerine dayalı bir araştırma gerçekleştirdi. Elde edilen sonuçlar neticesinde mevcut tüm özellikleri kullanmaktansa, KGI'nı etkili bir biçimde tahminini sağlamak için özellik seçim yöntemlerini kullanmanın önemini göstermiştir.

(Rao K ve ark., 2018) yaptıkları çalışmada, YSA algoritması ile geliştirilen tahmin modellerinde kullanılan 6 farklı girdi verisinde seçilen değişkenler birden altıya değiştirerek KGI'nın tahmin edilmesini amaçladılar. Her bir girdi verisinin farklı kombinasyonları ile girdi verisi temelinde 6 YSA modeli geliştirip (YSA-1'den YSA-6'ya) 32 olası girdi veri grubu için en iyi girdi kombinasyonunu belirlemek üzere seçilen verileri modeli eğitmek amacıyla kullandılar. Geliştirilen YSA modellerinden maksimum ve minimum sıcaklık farkı ve dünya dışı radyasyon girdi veri çiftinden oluşan tahmin modeli MAPE = %2.61 ile diğer 32 kombinasyondan daha iyi bir performans seviyesine ulaştığını ifade etmişlerdir.

(Khosravi ve ark., 2018b), İran'ın çeşitli iklim bölgelerinde bulunan 12 istasyonun meteorolojik ve coğrafik 11 verisi toplanarak günlük ortalama KGI'nı tahmin etmek amacıyla ML algoritmalarına dayalı farklı veri kombinasyonları seçilerek 20 farklı tahmin modeli geliştirdiler. Geliştirilen modellerde Grup Veri İşleme Yöntemi Sinir Ağı (GMDH), MLFFBPN, ANFIS, parçacık sürüsü optimizasyonu ile optimize edilmiş (ANFIS-PSO), genetik algoritma ile optimize edilmiş (ANFIS-GA) ve karınca kolonisi ile optimize edilmiş (ANFIS-ACO) algoritmalarını kullandılar. GMDH ve MLFFBPN ile geliştirilen modeller, diğer algoritmaları kullanılarak geliştirilen

modellerden daha iyi performans gösterdiğini belirttiler. Ayrıca, farklı iklim bölgelerindeki ANFIS-PSO modelinin, ANFIS-GA ve ANFIS-ACO modellerinden daha iyi bir performans seviyesine ulaştığını hesaplamışlardır.

(Akıl, 2019) yaptığı çalışmada, Güneydoğu Anadolu Bölgesinde bulunan beş ilin güneş ışınımını tahmin edebilmek amacıyla Lasso, Doğrusal, Bayesian Ridge, DVM Doğrusal, RF, Bagging, Ridge, Hubber, DVM RBF ve Adaboost regresyon algoritmaları olmak on farklı ML algoritması karşılaştırılmıştır. Girdi verilerinin hem tamamı hem de Temel Bileşen Analizi (PCA) sonrasında en etkili olanları seçilerek kullanılmıştır. PCA uygulandığında tahmin sonuçlarında belirgin iyileşme sağlandığı hesaplamıştır.

(Li ve ark., 2019) yaptıkları çalışmada, saatlik ortalama KGI'nı tahmin etmek amacıyla Multivariate Adaptive Regression Spline (MARS) ML algoritması ile yedi yıllık ölçülen saatlik veriler yardımıyla tahmin modelleri geliştirdiler ve Hong Kong şehrinde elde ettikleri sonuçları YSA ve Doğrusal Regresyon ML algoritmaları ile karşılaştırdılar. MARS algoritması kullanılarak en önemli 5 değişkeni belirlediler ve geliştirilen modellerde farklı kombinasyonlarını kullanarak 16 farklı girdi grubu oluşturdular. Yapılan karşılaştırmalı çalışmalarda YSA'nın diğer algoritmalara kıyasla daha üstün performansa ulaştığını belirttiler. Ancak, önerilen MARS modellerinin hem tahmin doğruluğu hem de yorumlana bilirlik açısından iyi bir performansa ulaştığını gösterdiler.

(El Mghouchi ve ark., 2019), günlük KGI'nın tahmin edilmesinde kullanılan veri seti değişkenlerinin en iyi girdi gruplarının belirlenmesinde YSA'na dayalı yeni bir yaklaşım sundular. 35 coğrafik yer için toplanan 15 farklı veri değişken için 2^{15}-32767 olası kombinasyon olduğundan ve bunun imkansızlığından hareketle, bu sayıyı düşürmek adına örtük değişkenleri çıkarıp seçilen 9 girdi değişkeni için 2^{9}-511 olası kombinasyonu hedef alan MLFFBPN mimarisi temelli tahmin modelleri geliştirdiler. Açıklık indeksi, dünya dışı radyasyon ve ortam sıcaklığı değişkenlerine dayanan kombinasyonların KGI'nı tahmin etmede oldukça başarılı olduğunu belirttiler.

(Ghimire ve ark., 2019) yaptıkları çalışmada, günlük ortalama KGI'nı tahmin etmek için Avustralya'nın Queensland bölgesinde bulunan 5 şehrin girdi verilerini kullandılar. Toplanan 85 farklı girdi değişkeninden her bir şehir için en uygun olanları belirlemede Komşuluk Bileşen Analizi (KBA) yönteminden faydalandılar. Veri odaklı yapılar olan DVR, Genetik Programlama (GP), Gaussian Süreci Makine Öğrenmesi (GPML) ve YSA ML algoritmalarını kullanarak geliştirilen modellerin performanslarını

değerlendirip, ARIMA, Sıcaklık Modeli (TM), Zaman serisi ve Fourier Serisi (TSFS) gibi istatistiksel modellerle karşılaştırmayı hedeflediler. YSA'nın performansının diğer ML algoritmalarından oldukça iyi olduğunu gösterdiler.

2.2. VM Yöntemleri Temelli Çalışmalar

VM birçok alanda farklı uygulaması olan bir yöntemdir. Özellikle tıptan mühendisliğe, sosyal bilimlerden bankacılık ve finansa kadar farklı alanlarda birçok çalışma yapılmıştır. Yapılan çalışmalardan bazıları aşağıda listelenmiştir.

(Alkan, 2014) yaptığı çalışmada, Kombinasyonel optimizasyon problemleri arasında zor sınıfında yer alan gezgin satıcı problemine kısa sürede iyi ve etkin bir çözüm bulunabilmesi amacıyla VM yöntemine dayanan bir yaklaşım geliştirmiştir.

(Gülen, 2014) yaptığı çalışmada, eğitimsel VM yöntemlerinden sınıflandırma analizi ile üstün yetenekli öğrencilerin ilgi alanlarının tahmin edilmesi ve birliktelik analizi ile bu öğrencilerin ilgi duydukları alanların belirlenmesini amaçlamıştır. Waikato Üniversitesinde geliştirilen Waikato Environment for Knowledge Analysis (WEKA) veri madenciliği yazılımı ile 10 sınıflandırma algoritması seçilmiş ve bu algoritmaların çıktıları karşılaştırılarak problem tanımı için en uygun olan algoritma JRip olarak belirlenmiştir.

(İnan, 2015), VM yöntemleri kullanılarak veri yığınlarından istenen bilgiye ulaşmak için kullanılması gerekmeyen veri özelliklerini tespit eden K-ortalamalar ve PCA algoritmalarına dayanan hibrid bir yöntem geliştirmiştir. Özellik çıkarım işlemi sonucunda YSA ve DVM sınıflandırıcılarının daha az nitelik ile çalıştırılması ile daha başarılı sonuçlara ulaşmıştır.

(Gültürk, 2016) yaptığı çalışmada, Cumhuriyet Üniversitesi Tıp Fakültesinde yatan kırım kongo kanamalı ateşi tedavisi gören hastaların 2009-2011 yılları arası verilerini kullanarak DVR, Random Forest ve Regresyon Ağacı ML algoritmaları ile R VM yazılımı kullanılarak kestirim performanslarının kıyaslanmasını amaçlamıştır. Senaryo yapılarına göre en iyi tahminin DVR ile hesaplandığını göstermiştir.

(Sağıroğlu, 2016) yaptığı çalışmada, VM bünyesinde seçilen algoritmalar ve paket programların değerlendirilmesi sonucunda yenilenebilir enerji yatırımlarına yönelik ihtiyaçlarla kaynaklar arasında maksimum faydaya yönelik modellemeler geliştirmiştir. En başarılı model Rapid Miner 5.1 programı kullanılarak Bulanık C-Ortalama (FCM) kümeleme algoritması ile alındığı belirtilmiştir.

(Cihan, 2018) yaptığı çalışmada, VM yöntemleriyle hayvan hastalıklarında teşhis, prognoz ve risk faktörlerinin belirlenmesi amaçlamıştır. Verilerin sınıflandırılması aşamasında, karar ağaçları, saf bayes, K-NN, YSA ve rastgele orman algoritmaları karşılaştırılmış ve en başarılı yöntemin saf bayes olduğu belirlenmiştir.

(Şahingil, 2019), birbirinden farklı genlere ait promoter bölgelerinin içerisinde bulunan nükleotit dizilimlerinin, gen promoter bölgeleri içerisindeki nükleotit bölgeleri ile genlerin fonksiyonalitesi ve sınıfları arasındaki ilişkileri, çeşitli veri madenciliği tekniklerini kullanılarak incelemiştir.

(Gülsoy, 2019), kobi-ticari-kurumsal müşterilerin kredilendirme sürecine yönelik VM tabanlı bir risk ölçme yöntemi geliştirmiştir. Çalışmasında bir bankanın kredi değerlendirme süreci üzerinde, mevcut müşteri verileri toplanarak 6 farklı VM sınıflandırma algoritmasından faydalanılmıştır. En başarılı sonuçlara %78.64 doğrulukla Multi Objective Evolutionary Fuzzy Classifier algoritması ile ulaşıldığı belirtilmiştir.

(Sevim, 2019) yaptığı çalışmada, VM yöntemiyle hastalıkların önceden tahmin edilmesi amacıyla diyabet hastalığı veri seti kullanılarak WEKA ve Python yazılımlarıyla çıkan sonuçlar karşılaştırılmıştır. Doğrusal ve Lojistik Regresyon algoritmaları kullanılarak geliştirilen tahmin modelleri için WEKA yazılımı ile daha kesin sonuçlara ulaştığı gözlemlenmiştir.

Veri madenciliği kullanılarak yapılmış olan yüksek lisans ve doktora tezleri incelendiğinde son zamanlarda oldukça fazla tercih edilen bir yöntem olduğu gözlemlenmiştir. Ancak, veri setindeki anlamsal bilgi ve özellikleri keşfetmeye odaklanan disiplinler arası bir yaklaşım olarak bilinen VM; YZ, deneysel ve istatistiksel yöntemler kadar güneş ışınım tahmininde doğrudan kullanılması pek fazla tercih edilen bir yöntem olmamıştır. Bununla birlikte, son zamanlarda VM ve ML yöntemlerinin aynı türden algoritmaları kullanarak birçok alanda yapılan çalışmaların artmasıyla kavramlar birbirlerinin yerine kullanılmaya başlamıştır. Bu sebepten, KGI tahmininde VM temelli yayınların sayısı armaya başlamıştır. Bununla birlikte, bazı çalışmalarda, modelin tahmin kesinliğini ve yorumlana bilirliğini arttırmak, modelleme süresini kısaltmak amacıyla VM süreçleri, ML yöntemleri ve YZ alanına giren tekniklerle beraber kullanılmıştır (Selbaş ve ark., 2011). VM temelli güneş ışınımın tahmin edilmesi ile ilgili yapılan çalışmalar aşağıda listelenmiştir.

(Mori ve Takahashi, 2012), KGI'nın tahmini edilmesinde VM süreçlerini kullanarak birden çok girdi değişkeni arasından sınıflama ve regresyon ağaçları tabanlı CART algoritması ile en uygun girdi verilerin seçimi için bir yöntem geliştirmişlerdir.

(Yadav ve ark., 2014) yaptıkları çalışmada, Hindistan'ın 26 adet bölgesine ait enlem, boylam, yükseklik, sıcaklık, en yüksek sıcaklık, en düşük sıcaklık ve güneşlenme süresi coğrafik ve meteorolojik verileri kullanılarak aylık ortalama günlük KGI'nı tahmin etmişlerdir. Veri madenciliği süreçlerini kullanarak güneş ışınımına en fazla etki eden girdi verilerini belirlemede WEKA programında bulunan J48 karar ağacı algoritmasından faydalanmışlardır. Girdi verileri kullanılarak üç farklı ANN modeli geliştirilmiştir. Birinci ANN modelinde girdi verilerinin tamamı, ikinci ANN modelinde veri madenciliği süreci sonunda seçilen 5 girdi verisi ve üçüncü ANN modelinde ise ikinci ANN modelinden belirlenen verilerden güneşlenme süreci çıkarılarak dört girdili bir model oluşturulmuştur. En iyi performans R = 0.9122 ve MAPE = 6.89 ile ikinci ANN modelinde elde edilmiştir. Bir başka çalışmada (Yadav ve ark., 2015), Hindistan'ın 76 bölgesinin aylık ortalama günlük güneş ışınım potansiyelini tahmin etmek amacıyla MLP, RBFNN, GRNN sinir ağı mimarilerinden faydalanarak karşılaştırmalı bir çalışma yapmışlardır. Dokuz farklı girdi parametresinden en etkili olanları belirlemede 2001 yılında YALE üniversitesinde geliştirilen Rapid Miner programından faydalanmışlardır. Program yardımıyla girdi değişkenlerinin sırasını veya ağırlığını belirlemek için PCA algoritması kullanılmış ve seçilen farklı girdi verileriyle her sinir ağı mimarisi için 5 farklı tahmin modeli geliştirilmiştir. En başarılı tahmin modeli sıcaklık, minimum sıcaklık, maksimum sıcaklık, yükseklik ve güneşlenme süresi girdi verilerinin kullanıldığı MLP sinir ağında MAPE = %6.89 değeri ile elde edilmiştir. (Yadav ve Chandel, 2015) bir diğer çalışmalarında, veri madenciliği süreçlerinden faydalanarak Hindistan'ın Himachal Pradesh eyaletinin güneş ışınımına en fazla etki eden optimum değişkenlerini belirlemiş ve güneş enerjisi potansiyelini YSA tabanlı ML algoritması ile tahmin etmişlerdir. WEKA programında J48 algoritmasının kullanılması ile 5 farklı YSA modeli geliştirilmiştir. Sıcaklık, maksimum sıcaklık, minimum sıcaklık, yükseklik ve güneşlenme süresi girdi veri seti ile geliştirilen model MAPE = %6.89 hesaplanan değer ile aylık ortalama günlük KGI'nı başarılı bir şekilde tahmin ettiğini belirtmişlerdir.

(Perez, 2016) yaptığı çalışmada, VM yöntemleri kullanarak güneş santrallerinin kısa vadede ne kadar enerji üreteceğini önceden bilmek amacıyla saatlik güneş ışınımını tahmin etmek için yeni modeller geliştirmiştir. Sınıflandırma tabanlı K-ortalamalar algoritmasını günün ait olduğu kümeyi belirlemek amacıyla, regresyon tabanlı Karar Ağaçları, YSA, DVM ve DVR algoritmaları ise günlük netlik endeksini tahmin etmek

amacıyla kullanılmıştır. Her iki yöntem ve önerilen ağırlık merkezleri sayesinde saatlik KGI'nı tahmin edilmesi sağlanmıştır.

(Abedin ve ark., 2017) yaptıkları çalışmada, VM ve YSA algoritmasını birlikte kullanarak Bangladeş'in farklı meteorolojik bölgelerinden toplanan verilerle KGI'nı tahmin etmeyi amaçlamışlardır. 12 farklı coğrafik ve meteorolojik veriden, WEKA programı yardımıyla, güneş ışınımına en fazla etki eden değişkenler belirlenmiştir. 7 girdili tahmin modeli R = 0.8306 MAPE = 0.092 ile en başarılı tahmin performansını göstermiştir.

(Ghofrani ve ark., 2017), güneş ışınımının tahmin kesinliğini arttırmak amacıyla VM destekli yeni bir model önerilmişlerdir. Bu modelde TS-Kümeleme algoritması ile sıcaklık, dünya dışı radyasyon ve rüzgâr hızı ve yönü verileri içinden en iyi veri kümesi bulunmuş ve RNN algoritmasıyla güneş ışınımı başarılı bir şekilde tahmin edilmiştir.

(Kayri ve ark., 2017) yaptıkları çalışmada, MLR, RF ve YSA gibi üç farklı ML algoritmasını kullanarak PV güneş gücünü tahmin etmede VM süreci destekli karşılaştırmalı bir çalışma amaçlamışlardır. Girdi olarak KGI, sıcaklık, rüzgâr hızı, rüzgâr yönü, bağıl nem ve güneş yükseklik açısı kullanmışlardır. Ayrıca modelleme yapmadan önce her bir algoritma ile duyarlılık analizi yaparak güce en fazla etki eden girdi parametrelerini de bulmuşlardır. Her üç yöntemde de güce en fazla etki eden girdi değişkeninin KGI olduğunu gözlemlemişlerdir.

(Meenal ve Selvakumar, 2018) yaptıkları çalışmada, model hatasını artırmadan KGI'nın kolay tahmin edilmesi için özellik sayısını azaltmayı ve en etkili girdi değişkenlerini kullanarak DVM, YSA ve deneysel modellerin tahmin doğruluğunu iyileştirmeyi amaçlamışlardır. Veri madenciliği süreçlerini kullanarak ay, enlem, maksimum sıcaklık ve güneşlenme süresi verilerinin güneş ışınımı üzerinde en fazla etkiye sahip olduğunu bulmuşlardır. Özellik seçiminde WEKA programının Greedy Stepwise özellik seçim yönteminden ve tahmin modellerinde kullanılan algoritmaların duyarlılık analizinden faydalanmışlardır. Ardından seçilen girdi verileri grupları ile birden fazla tahmin modelleri geliştirip karşılaştırmışlardır. WEKA yazılımında kullanılan DVM tabanlı ML algoritması ile geliştirilen modelin, YSA ve deneysel yöntemlerden daha iyi tahmin kesinliğine ulaştığını hesaplamışlardır.

(Lotfinejad ve ark., 2018), İran'ın günlük KGI potansiyelinin tahmin edilmesi amacıyla yarasa algoritması ve sinir ağlarının belirli bir kombinasyonuna dayanan Yarasa Sinir Ağı (BNN) olarak tanımlanan yeni çok katmanlı hibrit bir model geliştirildiler. Geliştirilen modeli GRNN ve ANFİS tahmin modelleri ile karşılaştırmayı

amaçladılar. İran'ın dört şehri için toplanan 10 meteorolojik veriden VM tabanlı PCA algoritmasını kullanarak optimum beş girdi değişkeni belirlediler. BNN tahmin modelinin tüm şehirlerde KGI'nı başarılı bir şekilde tahmin ettiğini gösterdiler.

(del Campo-Ávila ve ark., 2019) yaptıkları çalışmada, güneş spektrumu ışınım dağılımını tahmin etmek amacıyla çeşitli VM yöntemlerine dayanan bir sistem geliştirmişlerdir. WEKA yazılımı kullanılarak RF algoritması ile meteorolojik parametreye karşılık gelen spektrum türleri belirlenir, M5P algoritması ile meteorolojik değişkenlere karşılık gelen güneş spektrumu tahmin edilmektedir. Önerilen yöntemin herhangi bir konumda uygulanabileceği belirtilmiştir.

(Al-Rousan ve ark., 2021), Ürdün'ün saatlik güneş ışınımını tahmin etmede VM yöntemleri içinde gruplandırılan Kural, Ağaç, Meta, Tembel ve Fonksiyon tabanlı 5 farklı yöntemden 16 farklı algoritmayı belirleyip modeller geliştirerek birbirleri ile karşılaştırılmayı hedeflemişlerdir. Her grup içinde belirlenen algoritmalar birbirleri ile karşılaştırıldığında sırasıyla M5, RF, Bagging, IBK ve MLP algoritmaları ile geliştirilen tahmin modelleri daha üstün başarı göstermiştir. Bu beş algoritma birbirleri ile karşılaştırıldığında, RF algoritması ile geliştirilen model R^2 = 0.9637 ile KGI'nı diğer algoritmalardan daha iyi tahmin ettiğini hesaplamışlardır.

Yapılan kapsamlı literatür araştırmalarında, VM süreçlerinin birçok alanda kullanılan yöntemlerin başında gelmesine rağmen, KGI'nın tahmin edilmesinde oldukça az yararlanıldığı gözlemlenmiştir. Ayrıca, yapılan çalışmalarda algoritma-tahmin ilişkisinin ön plana alınması ile veri üzerinde yapılacak ön işlem süreçlerine çok fazla zaman ayrılmadığı anlaşılmaktadır. Son zamanlarda, veri-tahmin ilişkisini hedef alan deneme-yanılma veya farklı yöntemlerle veri seti üzerinde birtakım işlemler yapılsa da bunların kapsamının oldukça sınırlı olduğu gözlemlenmiştir. Bununla birlikte, literatür taramalarında veri setinin yaygın olarak günlük ve aylık ölçekte seçilip kullanıldığı, saatlik ölçekte yapılan çalışmaların çok az olduğu anlaşılmaktadır. Ayrıca Türkiye'nin birçok yeri ile ilgili kapsamı sınırlı olan YZ tabanlı tahmin çalışmaları olmasına rağmen, Akdeniz Bölgesinin tamamını hedef alan, saatlik ölçekte ve kapsamlı bir çalışmaya rastlanılmadığından, hedeflenen çalışmanın bölge içinde belirlenen sekiz ilin verileri kullanılarak yapılması kararlaştırılmıştır.

3. MATERYAL VE YÖNTEM

Bu bölümde, Türkiye'nin güneş enerjisi potansiyelinin durumu ve çalışma alanı olarak belirlenen Akdeniz bölgesinin iklimsel, coğrafik ve güneş ışınım özelliklerine ayrıntılı olarak değinilmiştir. Ayrıca bölge içinde seçilen sekiz ilin tahmin modellerinde yararlanılacak güneş tabanlı verilerinin yapısı ve seçim süreçleri anlatılmıştır. Geliştirilen tahmin modellerinde faydalanılan ML algoritmaların yapısal özellikleri, modelleme süreçleri ve model performanslarının değerlendirilmesinde kullanılan istatistiksel ölçekler ve analiz yöntemleri ele alınmıştır.

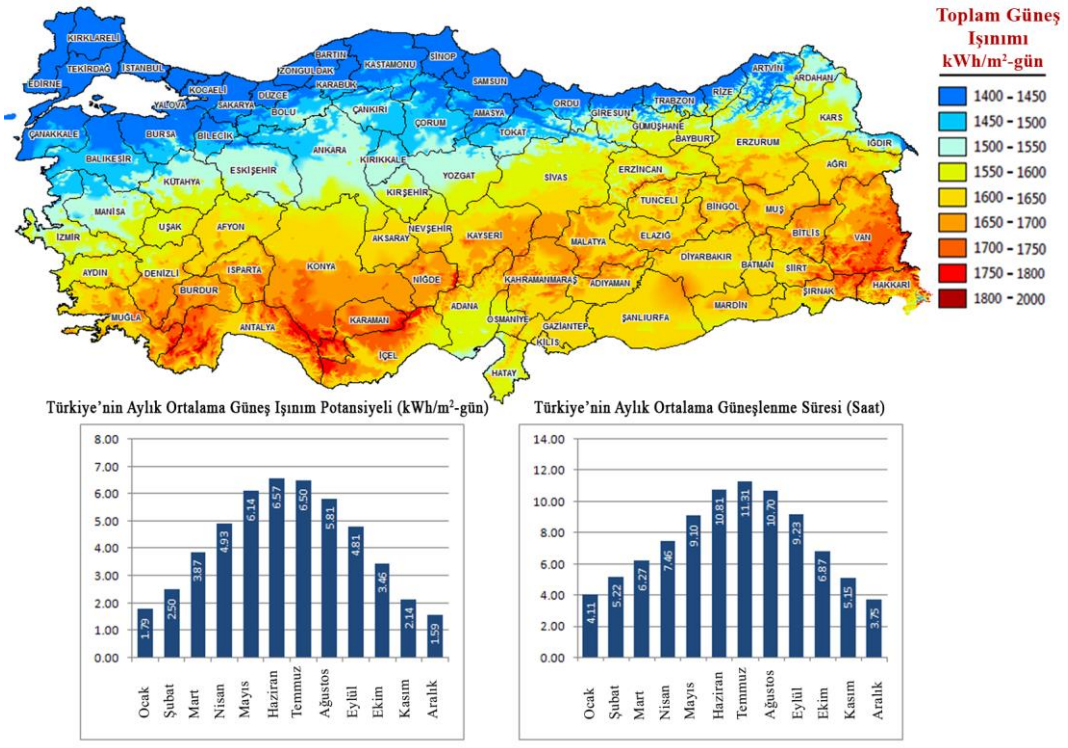
3.1. Türkiye'de Güneş Işınım Potansiyeli

Türkiye Asya ve Avrupa kıtalarını birleştiren, 36° – 42° kuzey paralelleri ile 26° – 45° doğu meridyenleri arasında kalan bir orta kuşak ülkesi olarak bilinmektedir. Ülke yedi farklı bölgeye ayrılmış ve Akdeniz güneş kuşağı ülkenin içinden geçerek Türkiye'yi güneş enerjisi potansiyeli olarak Avrupa'nın en avantajlı yerlerinden birisi haline getirmiştir. Bu sebeple güneş ışınım potansiyeli, Avrupa'nın en yüksek güneş ışınım değerlerine sahip olan İspanya ve Portekiz'e benzemektedir. Ayrıca diğer Avrupa ülkeleri ile karşılaştırıldığında güneş enerjisi imkanları oldukça yüksek bir seviyede bulunmaktadır (Cengiz ve Mamis, 2015).

Türkiye Güneş Enerjisi Potansiyeli Atlasına (GEPA) göre ortalama yıllık toplam güneşlenme süresi 2,741 saat, ortalama günlük toplam 7.5 saat, ortalama yıllık KGI potansiyeli 1,527 kWh/m^2 olarak hesaplanmıştır. Ülkenin ortalama yıllık 114 gün gibi muazzam bir güneş enerji potansiyeli bulunmaktadır. Şekil 3.1'deki GEPA atlasında, 500x500 metre çözünürlükte 3 milyonun üstünde grid oluşturulmuş ve her bir grid için güneş kaynak bilgileri hesaplanarak coğrafi bilgi sistemleri üzerinden haritalama işlemleri gerçekleştirilmiştir. Geliştirilen modelin doğruluğunun hesaplanması ve kalibrasyonunun yapılması amacıyla 1985-2006 yıllarına ait ölçüm yapan 156 adet meteorolojik istasyonlardan elde edilen saatlik KGI verileri kullanılmıştır. GEPA haritasından, ülkenin tamamı veya herhangi bir şehrinin güneş ışınım değerleri ve sürelerinin aylık ortalamaları değerleri hakkında bir fikir edinilebilmektedir (EİGM, 2016). Türkiye de en sıcak aylar Temmuz ve Ağustos, en soğuk aylar Ocak ve Şubat olmakla birlikte kuzeyden güneye doğru inildikçe güneşlenme süresi önemli ölçüde artmaktadır (Uyan, 2017). Çizelge 3.1'de Elektrik İşleri Etüt İdaresi Genel Müdürlüğü

(EIE) tarafından 2006 yılında hazırlanan raporda Türkiye'nin güneş ışınımı potansiyelinin bölgelere göre dağılımı gösterilmiştir (TMMOB, 2018). Çizelgeye göre en az ışınım potansiyeline sahip olan bölgemiz Karadeniz Bölgesi olarak hesaplanmıştır. Diğer bölgelerimiz ise orta ve yüksek seviyede ışınım değerlerine sahiptir. Çizelgeden en fazla ışınım potansiyeline sahip olan bölgenin, Akdeniz ve Güneydoğu Anadolu bölgesi olduğu görülmektedir. Ortalama günlük güneşlenme süresi ve güneşli gün sayısı diğer bölgelere göre oldukça yüksek bulunmuştur. Bu nedenle güneş enerjisine dayalı yatırım yapmak amacıyla, bu iki bölgedeki coğrafik bir yerin seçimi, üretim potansiyeli açısından daha uygun olacağı anlaşılmaktadır.

Türkiye genel olarak güneş enerji potansiyeli yüksek bir ülkedir. Yıllık güneş ışınım potansiyeli 1,200 kWh/m^2 seviyelerinde olan ve güneş enerjisine oldukça fazla yatırım yapan ülkelerin başında gelen Almanya ile karşılaştırıldığında, Türkiye'nin en az ışınım potansiyeline sahip olan bölgeleri Karadeniz ve Marmara'ya benzer olduğu görülmektedir (TMMOB, 2018). Ülkemizden oldukça düşük güneş potansiyellerine sahip olan ülkelerin güneş enerjisine ayrı bir parantez açmaları, ülkemizin bu alanda yapılacak yatırımların önemini bir kez daha ortaya koymaktadır.



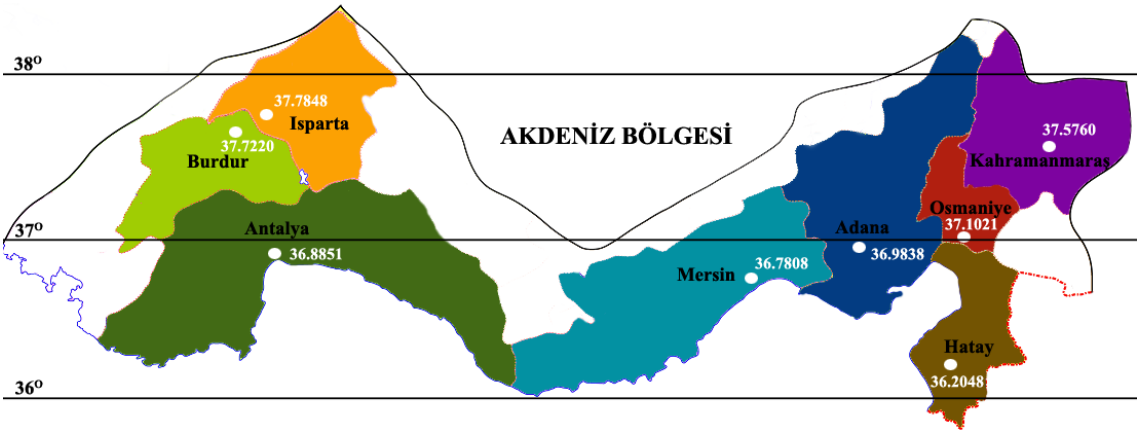
Şekil 3.1. Türkiye'nin güneş enerjisi potansiyeli atlası (GEPA) (EİGM, 2016)

Çizelge 3.1. Türkiye'nin bölgelere göre güneş ışınım potansiyelleri

Türkiye'nin Bölgeleri	Ort. Toplam Güneş Işınım Şiddeti (kWh/m^2 - Yıl)	Ort. Güneşlenme Süresi (Saat/Yıl)	Ort. Güneşlenme Süresi (Saat/Gün)	Ort. Güneşli Gün Sayısı (Gün)
Güneydoğu Anadolu	1,460	2,993	8.2	125
Akdeniz	1,390	2,956	8.1	123
Doğu Anadolu	1,365	2,664	7.3	111
İç Anadolu	1,314	2,628	7.2	110
Ege	1,304	2,738	7.5	114
Marmara	1,168	2,409	6.6	100
Karadeniz	1,120	1,971	5.4	82

3.2. Çalışma Alanı ve Verilerin Seçimi

Akdeniz bölgesi, yaklaşık olarak 120,000 km^2'lik yüzölçümüyle Türkiye'nin %15'ni kapsamaktadır. En sıcak ay ortalaması 26 ile 28 °C arasında, en soğuk ay ortalaması ise 8 ile 10 °C arasında olmaktadır. Yıllık sıcaklık ortalaması ise yaklaşık 18 °C civarındadır. Bölge kıyıları kış mevsiminin en ılık geçtiği yerler olarak bilinmektedir (Doğaner, 2020). Çizelge 3.1'de görüleceği üzere, bölgede 1,390 kWh/m^2 gibi ortalama yıllık güneş ışınım şiddeti ve 2,956 saat gibi ortalama yıllık güneşlenme süresi hesaplanmıştır. Ayrıca, (Yurtsever ve Öztopal, 2017) yaptıkları bir çalışmada, Akdeniz bölgesinin 1999 - 2015 arası uydu görüntüleri ile GEPA atlasındaki verileri karşılaştırmışlardır. Uydu ölçüm verilerine göre Akdeniz bölgesinin ışınım değerlerinin yıllık ortalamalarını 1,751 ile 2,014 kWh/m^2 aralığında, en düşük ise 1,576 kWh/m^2 seviyelerine hesaplamışlardır. Akdeniz bölgesinin oldukça yüksek seviyelerde güneş enerjisi potansiyeline sahip olduğunu ve en sağlıklı sonuçların yer istasyonlarının gözlemleri ile alınacağını ifade etmişlerdir. Bölge üzerinde ışınım temelli yapılan çalışmalar, meteorolojik veriler ve bölgenin konumu sebebiyle yüksek bir güneş enerji potansiyeline sahip olduğu anlaşılmıştır. Bu sebeple yapılacak tez çalışmasının Akdeniz bölgesi temelli olmasına karar verilmiştir. Bölgede var olan, Adana, Antalya, Burdur, Hatay, Isparta, Kahramanmaraş. Mersin ve Osmaniye illerinin tamamı çalışma alanı olarak belirlenmiştir. Şekil 3.2'de seçilen illerin Akdeniz bölgesi içindeki konumları ve meteorolojik ölçüm istasyonlarının enlem koordinatları verilmiştir.



Şekil 3.2. Akdeniz bölgesinde seçilen iller ve meteorolojik ölçüm istasyonlarının enlem koordinatları

Güneş enerjisi tabanlı sistemlerinde, ışınım verisi, kurulacak sistemin planlanıp tasarlanmasında ve üretilecek enerjinin öngörülüp kazanç durumunun belirlenmesinde en önemli parametre olarak ölçülmesi gerekmektedir. Ancak birçok ülkede olduğu gibi ülkemizde de her meteorolojik ölçüm istasyonunda ışınım verisi ölçülememekte veya sınırlı sayıda ölçülebilmektedir. Ülkemizde, Meteoroloji Genel Müdürlüğü (MGM), hava tahminleri ve gözlemler yapmak amacıyla ülkenin birçok yerinde farklı amaçlarla kullanılan meteorolojik ölçüm istasyonlarını açmak ve çalıştırmakla yükümlü olarak kurulmuştur. Ancak ülkenin tamamına yayılan ve geniş bir istasyon ağı olan bu kurumda bile sadece belirli lokasyonlarda güneş ışınım verisi sağlıklı bir şekilde ölçülebilmektedir. KGI tahmin çalışmasında belirlenen iller için yararlanılacak veriler MGM'den talep edilmiştir. İlk etapta 2013 – 2018 yılları arasında 5 yıllık saatlik ölçülmüş veriler istenmiş, ancak büyük oranda eksikliklerle karşılaşılınca yapılan yazışmalar ve yerinde yapılan görüşmeler neticesinde 2002 – 2006 arasındaki verilerin kullanılmasına karar verilmiştir. Bu kararın alınmasındaki en önemli sebep 2007'den sonra güneş ışınımını ölçen snoptik istasyonların hedeflenen bazı yerlerde ölçümlerinin durdurularak, otomatik istasyonlara geçilmesi olmuştur. Otomatik istasyonların bir kısmında da güneş ışınım verileri ölçülmüş, ancak çok fazla eksiklikler gözlemlendiğinden çalışmanın bölgenin genelini kapsayamayacağı düşüncesiyle vazgeçilmiştir.

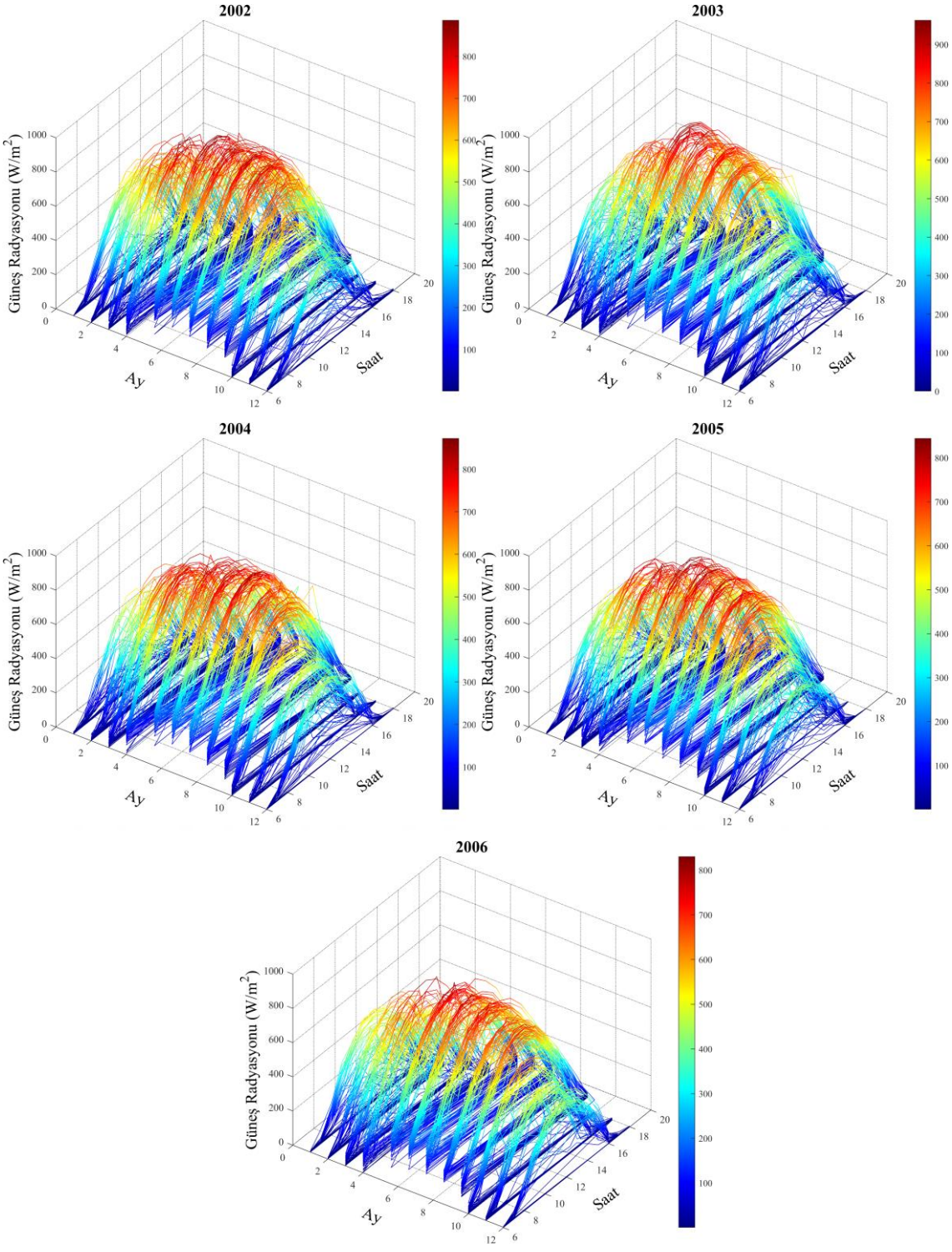
MGM'den istenen veriler her 5 dakikada bir ölçülmüş saatlik ortalama verilerdir ve belirlenen her bir ile ait ölçüm istasyonlarından Basınç (B), Sıcaklık (S), Nem (N), Rüzgâr Hızı (RH), Güneşlenme Süresi (GS) ve Küresel Güneş Işınımı (KGI) meteorolojik talep edilmiştir. Bu veri setinin seçilmesinde yapılan önceki çalışmalar

etkili olmuştur. Çizelge 3.2'de toplanan verilerin özellikleri ve dağılımları gösterilmiştir. Verilerin hepsi saatlik olarak ölçülmüştür. Her bir ilin verilerinin alındığı meteorolojik ölçüm istasyonlarının coğrafik koordinatları çizelgede gösterilmiştir.

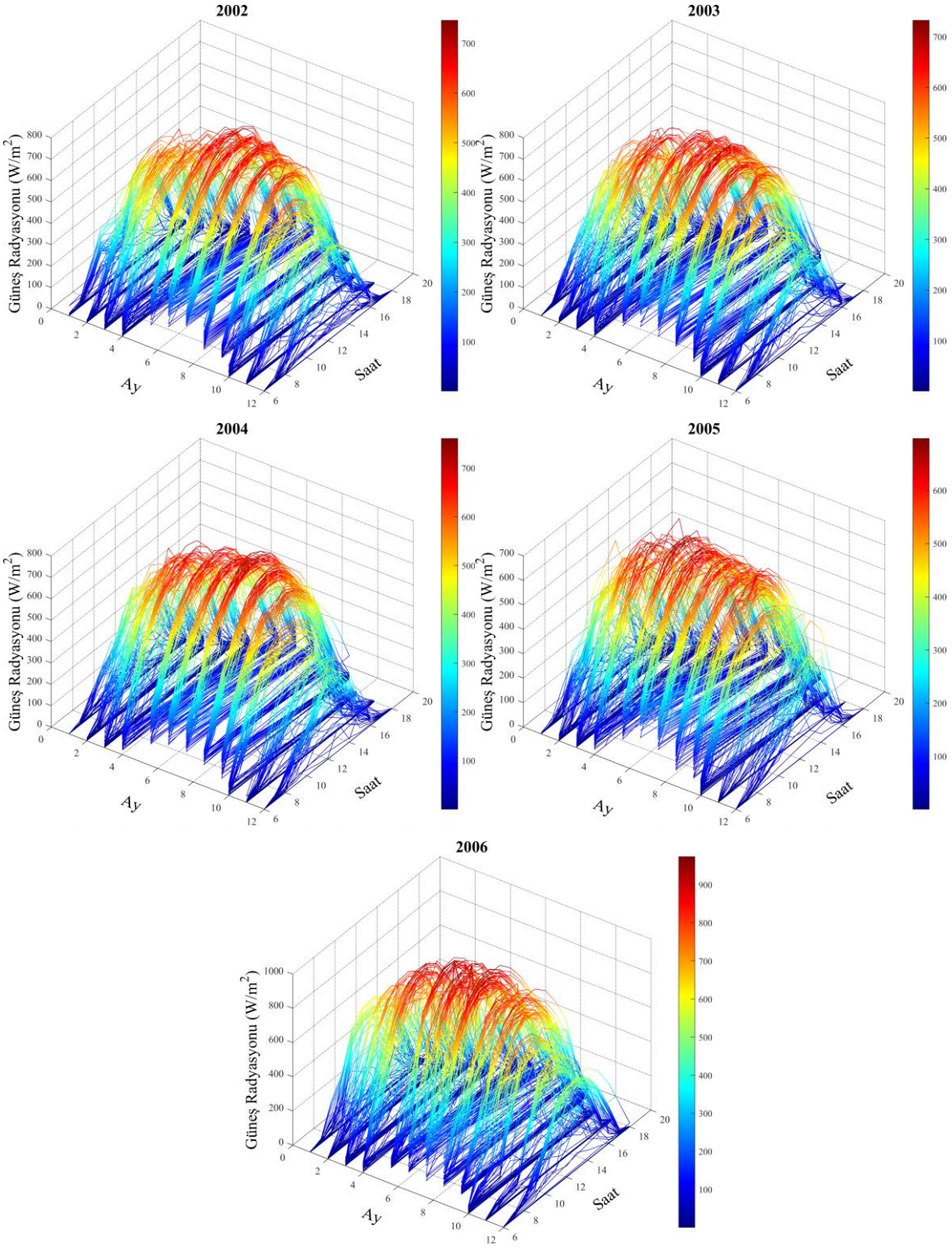
Çizelge 3.2. Çalışmada seçilen her bir ilin coğrafik ve meteorolojik verilerinin spesifik özellikleri

İller	Coğrafik Veriler			Meteorolojik veriler											
	Enlem	Boylam	Yük.	S (°C)		B (hPa)		N (%)		RH (m/s)		GS (Saat)		KGI (W/m^2)	
				Min	Mak	Min	Mak	Min	Mak	Min	Mak	Min	Mak	Min	Mak
Adana	37.041	35.344	23	-3	41	993.0	1028.2	6	99	0	9.9	0	1	0.0	963.0
Antalya	36.885	30.683	47	-3	42.1	973.8	1025.0	7	100	0	9.8	0	1	0.1	973.9
Burdur	37.722	30.294	957	-13	40.3	881.8	922.2	8	100	0	9.8	0	1	0.1	1053.7
Hatay	36.205	36.151	104	-3	41	952.5	1031.5	1	100	0	7.6	0	1	0.2	802.5
Isparta	37.785	30.568	1050	-14	36.8	876.0	918.9	7	99	0	10.1	0	1	0.1	1046.7
K.Maraş	37.576	36.915	872	-8	43.9	928.3	966.1	0	99	0	9.9	0	1	0.2	1028.8
Mersin	36.781	34.603	7	0.4	37.9	990.6	1031.8	10	96	0	9.9	0	1	0.1	949.0
Osmaniye	37.102	36.254	94	-1	43.1	978.2	1018.5	7	100	0	9.2	0	1	0.0	914.1

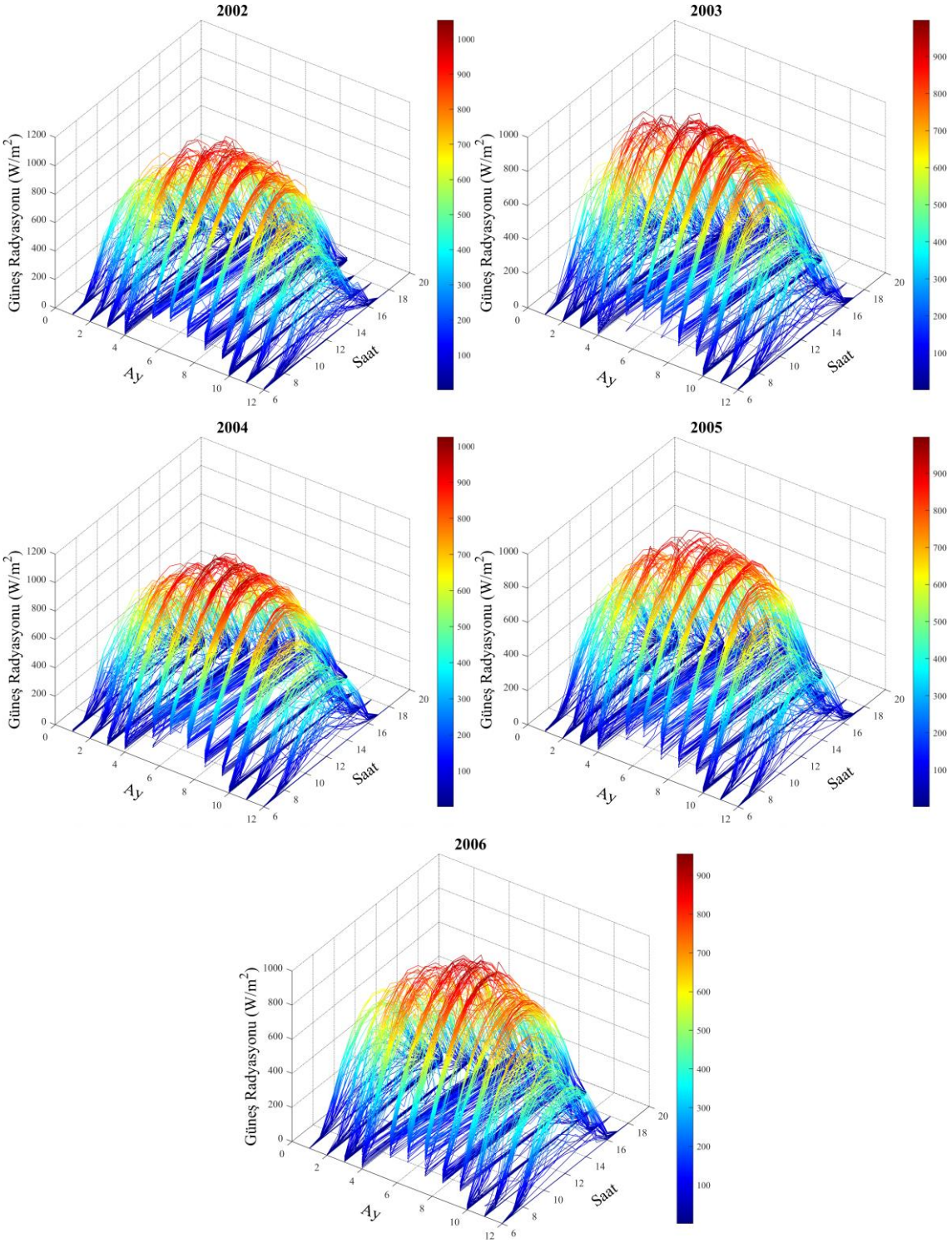
Akdeniz bölgesindeki 8 ilin 2002-2006 yılları arasında ölçülen KGI verisine göre hem aylık hem de mevsimsel olarak değişimini gösterir 3 boyutlu grafikler Şekil 3.3 – 3.10'da verilmiştir. Bu grafiklerde ölçülen güneş ışınım değerlerinin saatlik bazda yıllık dağılımları ayrıntılı olarak gösterilmiştir. Her bir ilin toplanan KGI verilerinin karakteristik dağılımlarından en az ve en çok etkili olduğu dönemler anlaşılabilmektedir. Ayrıca bu grafiklerde, seçilen verilerin ölçüm zaman aralıkları ve her bir ilin yıllık güneş enerjisi potansiyellerinin grafiksel gösterimi verilmiştir.



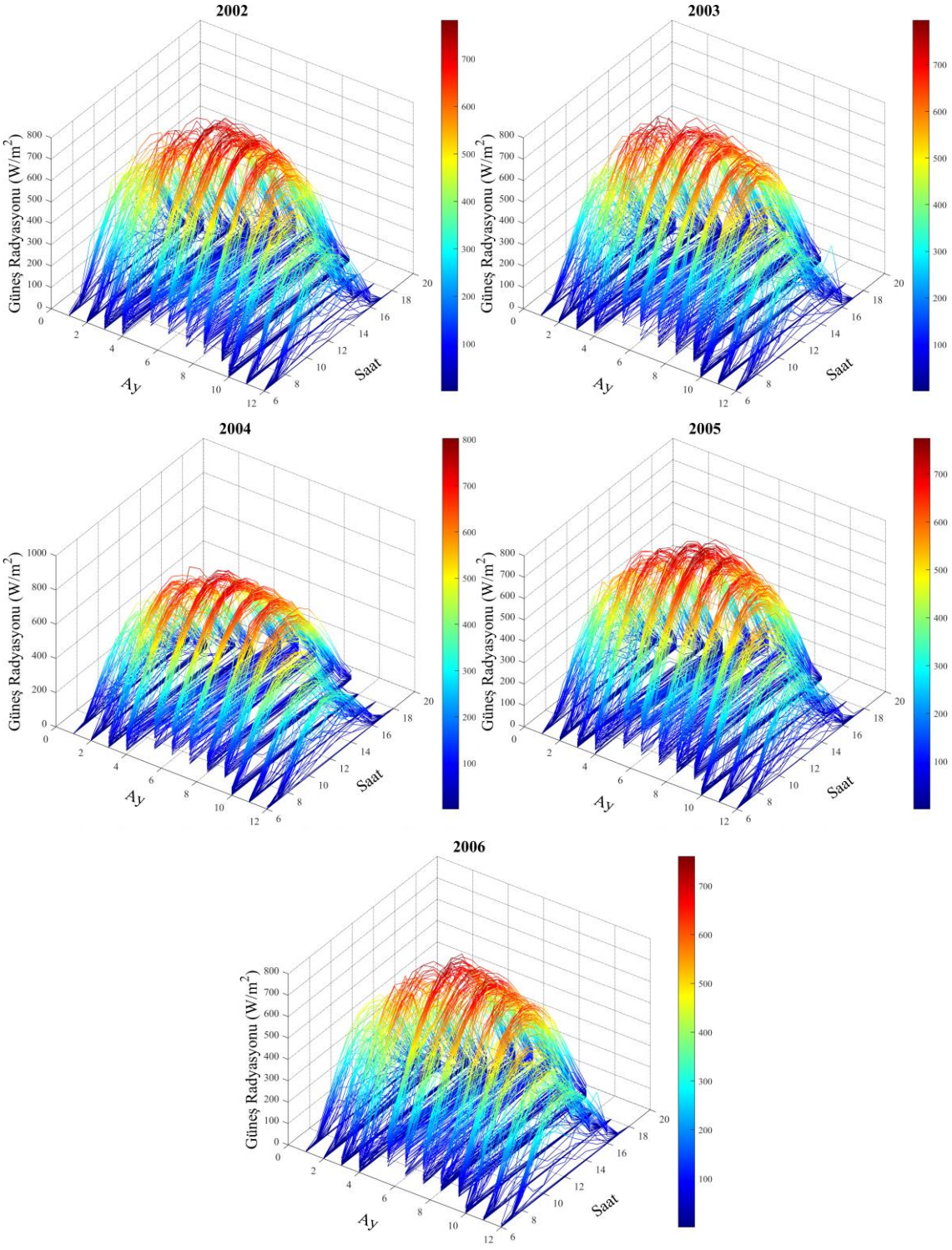
Şekil 3.3. Adana ilinin 2002-2006 yılları arasında ölçülen KGI'nın 3-Boyutlu yıllık dağılım grafikleri



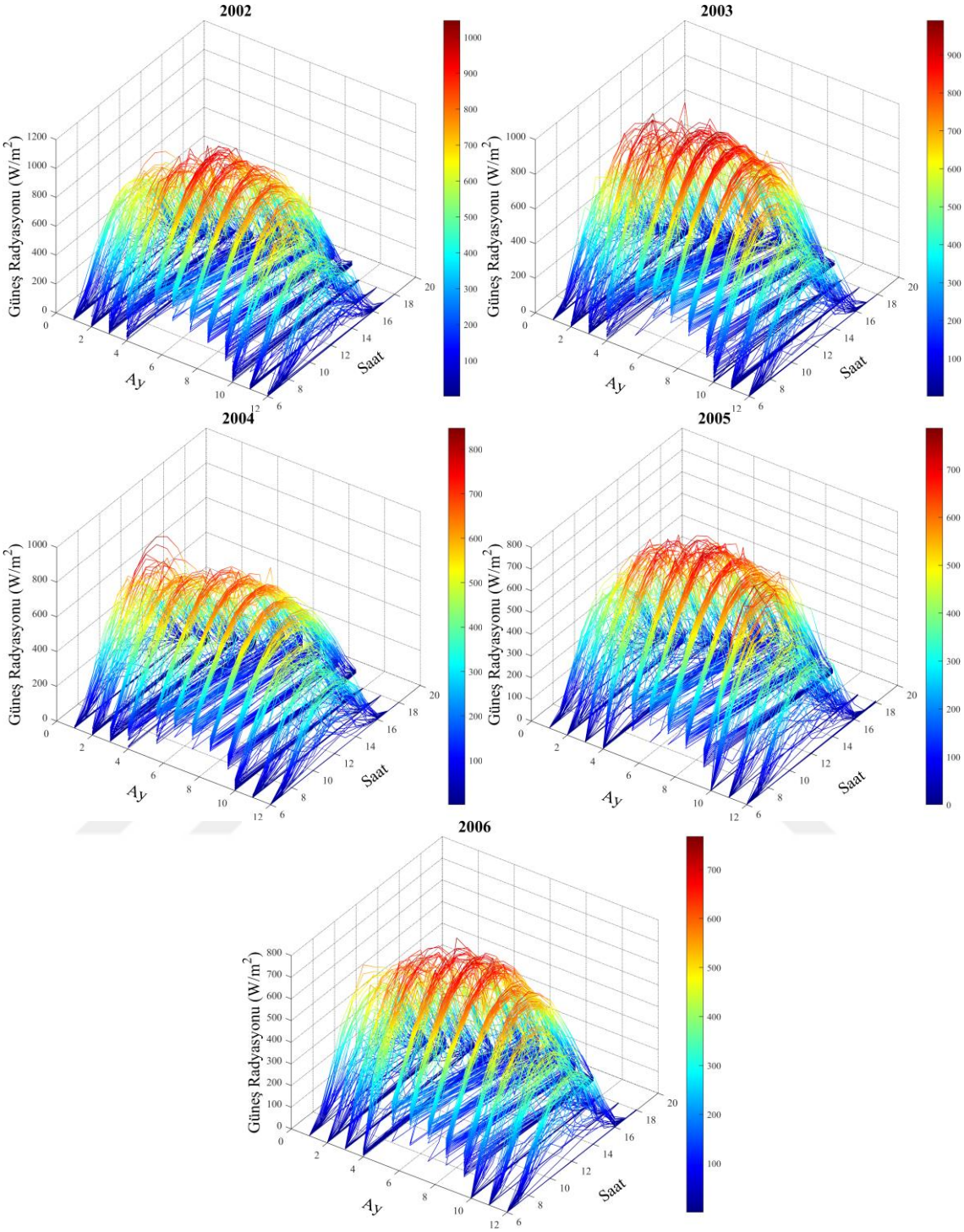
Şekil 3.4. Antalya ilinin 2002-2006 yılları arasında ölçülen KGI'nın 3-Boyutlu yıllık dağılım grafikleri



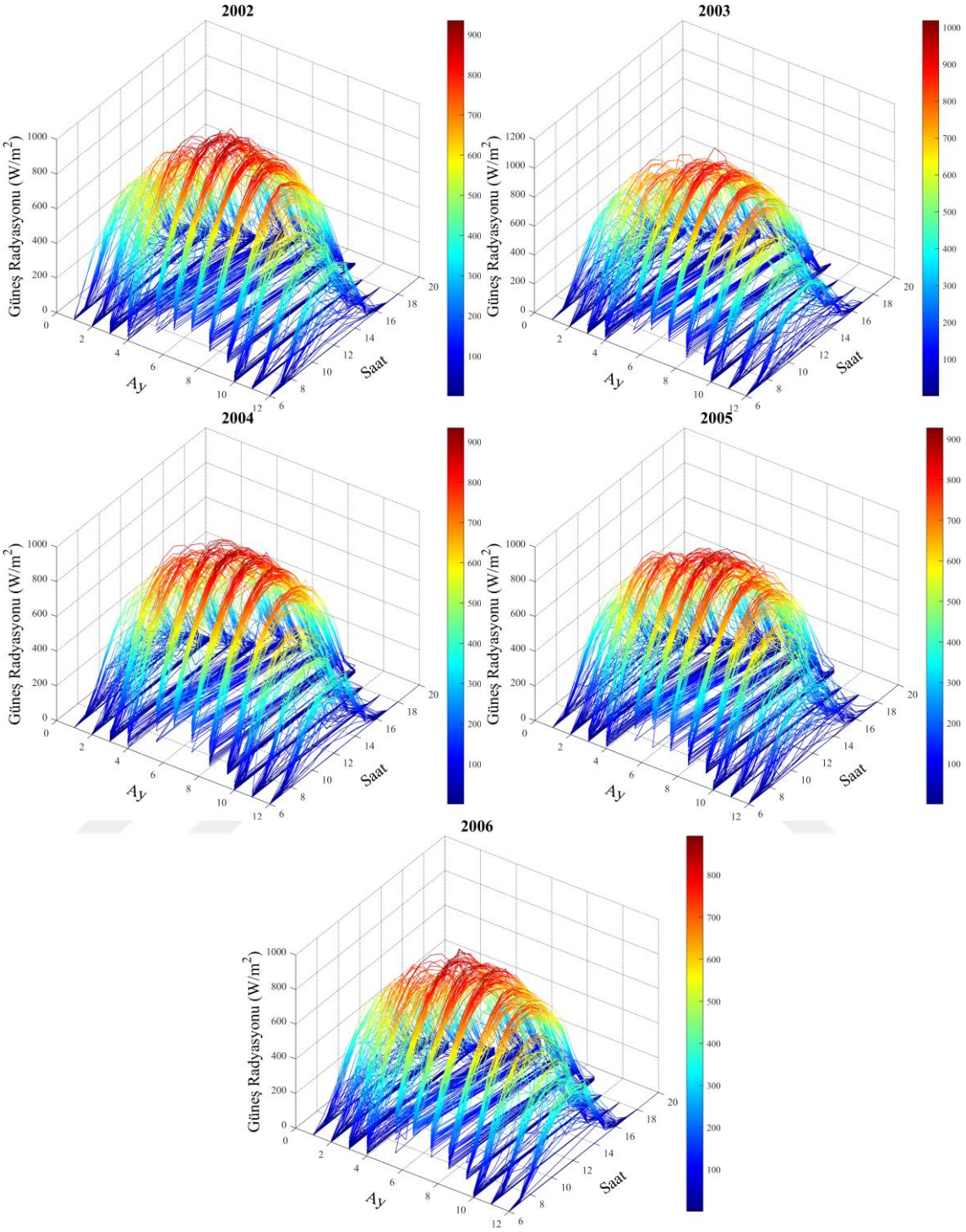
Şekil 3.5. Burdur ilinin 2002-2006 yılları arasında ölçülen KGI'nın 3-Boyutlu yıllık dağılım grafikleri



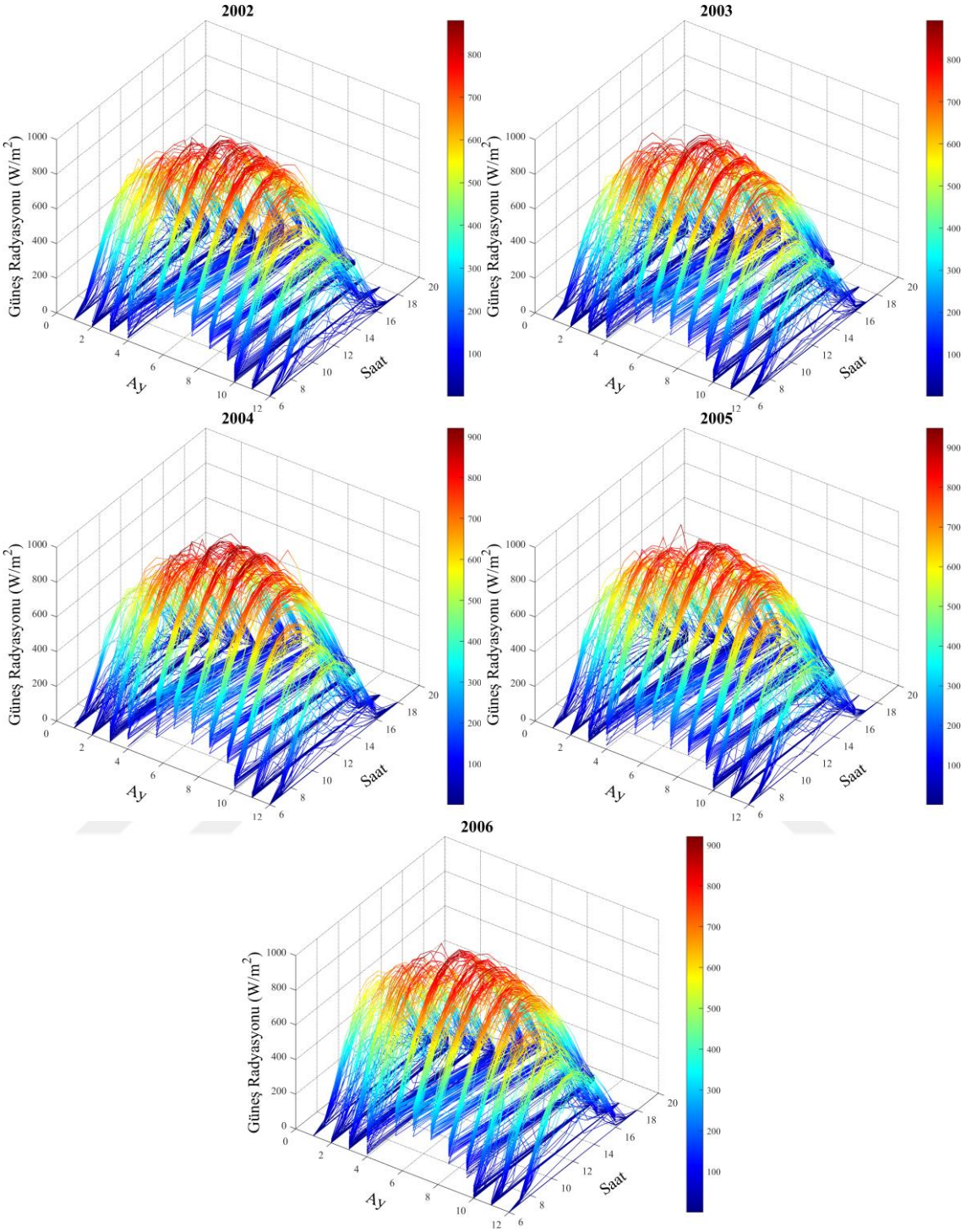
Şekil 3.6. Hatay ilinin 2002-2006 yılları arasında ölçülen KGI'nın 3-Boyutlu yıllık dağılım grafikleri



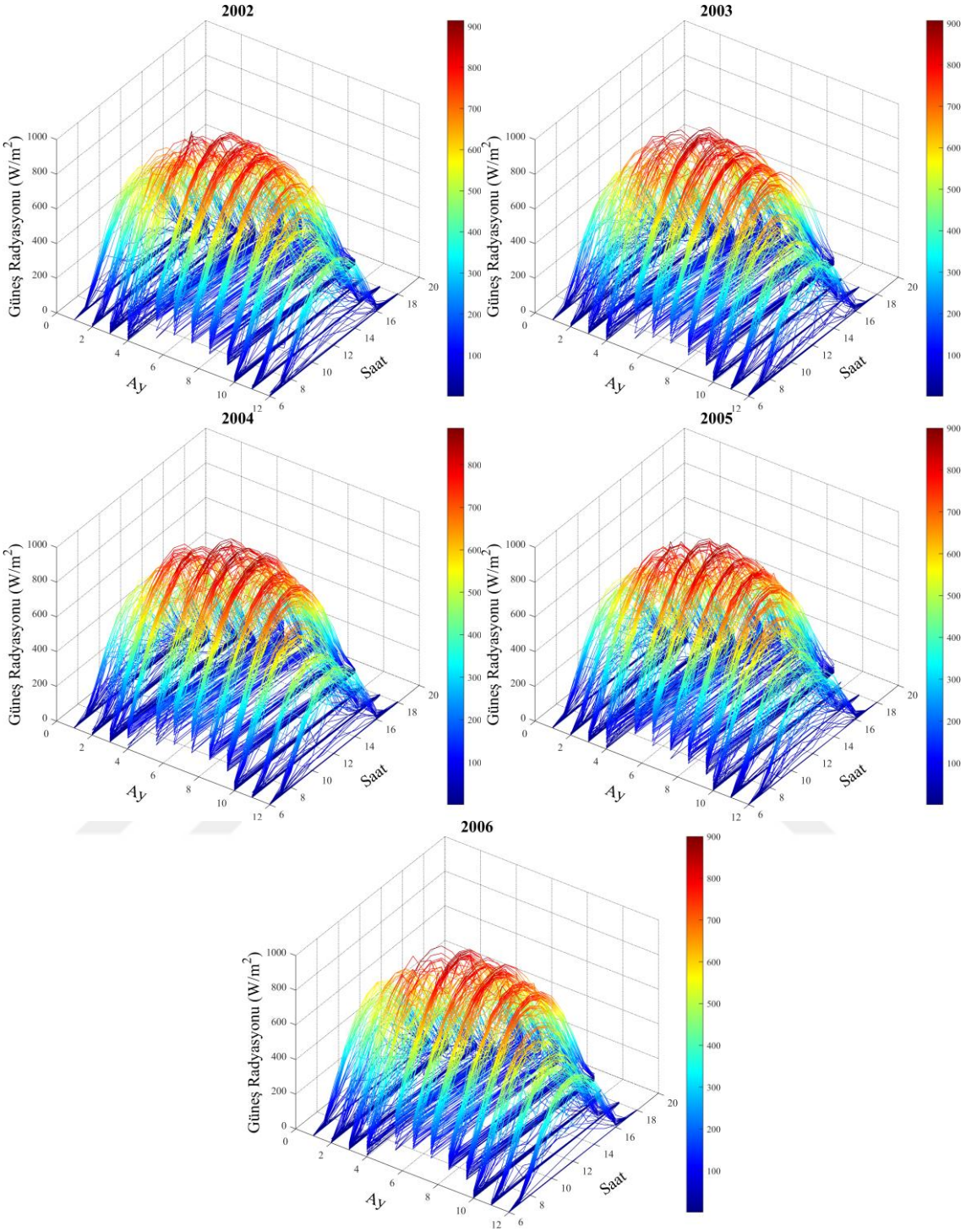
Şekil 3.7. Isparta ilinin 2002-2006 yılları arasında ölçülen KGI'nın 3-Boyutlu yıllık dağılım grafikleri



Şekil 3.8. Kahramanmaraş ilinin 2002-2006 yılları arasında ölçülen KGI'nın 3-Boyutlu yıllık dağılım grafikleri



Şekil 3.9. Mersin ilinin 2002-2006 yılları arasında ölçülen KGI'nın 3-Boyutlu yıllık dağılım grafikleri



Şekil 3.10. Osmaniye ilinin 2002-2006 yılları arasında ölçülen KGI'nın 3-Boyutlu yıllık dağılım grafikleri

3.3. ML Tabanlı Tahmin Algoritmaları

Makine Öğrenmesi, herhangi bir konuda karşılaşılan soruna çözüm bulmak amacıyla ilgili verilerden faydalanılarak modellerin geliştirilmesini veya öğrenmenin sağlanmasını amaçlayan algoritmalar bütünüdür. Seçilen bir algoritma farklı tiplerde

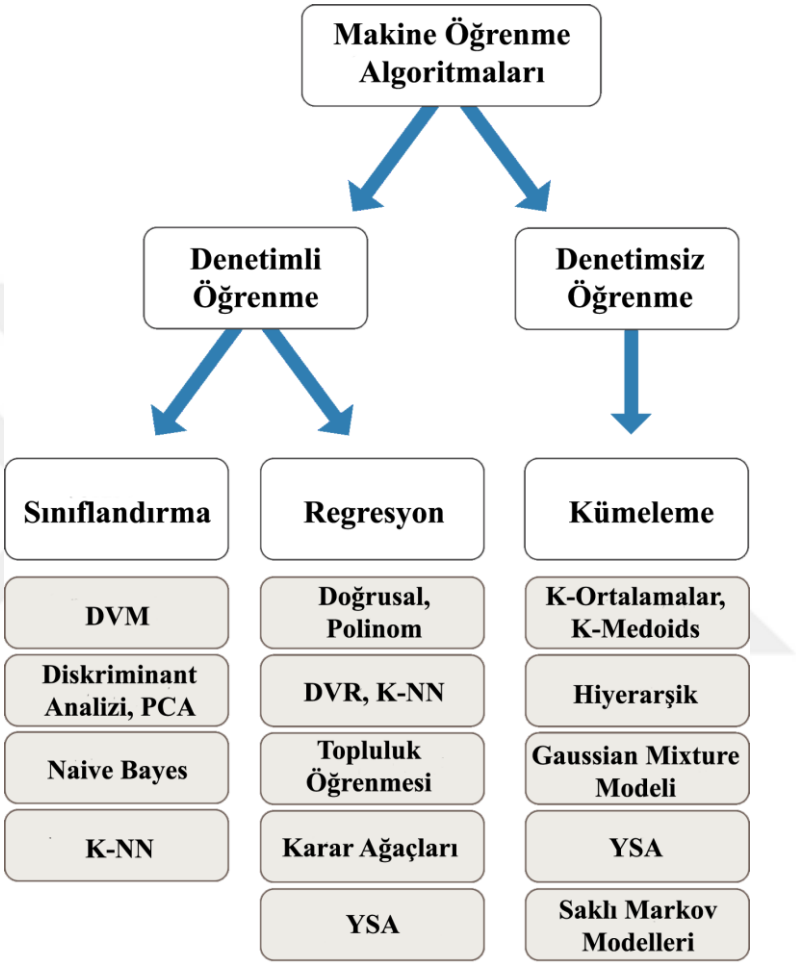
problemlerin çözümünde kullanılabilmektedir. Problemin türüne göre öğrenmenin oluşmasını sağlayan görevleri, Denetimli (Supervised), Denetimsiz (Unsupervised), Yarı Denetimli (Semi-Supervised) ve Takviyeli (Reinforcement) öğrenme tekniklerine dayanan ML algoritmaları ile gerçekleştirilmektedir (Sammut ve Webb, 2010).

Denetimli öğrenme, eğitim verileri kullanılarak geliştirilen modelin bir çıkış fonksiyonu üretmesini sağlayan öğrenme tekniği olarak bilinmektedir. Eğitim verileri girdi ve çıktı olarak tanımlanan etiketlenmiş verilerinden oluşmaktadır. Kullanılan algoritmalar, girdilerle hedef çıktılar arasında eşleme yapan bir fonksiyon üreterek öğrenme işlemini gerçekleştirirler. Bir başka ifadeyle bir veri setinin hangi sınıfa ait olduğunun bilindiği durumlarda bu türden algoritmalar kullanılmaktadır. Algoritmanın öğrenme süreci, kabul edilebilir bir performans seviyesine ulaşıldığında durmaktadır. Bu öğrenme yöntemi, sınıflandırma (kategori tahmini) ve regresyon-eğri uydurma (sayı tahmini) problemleri olarak iki ayrı kategoriye ayrılmaktadır (Sammut ve Webb, 2010).

Denetimsiz öğrenme, eğitim veri setinin sadece girdilerden oluştuğu, girdilere karşılık çıktıların olmadığı problemlerin çözümünde uygulanan bir öğrenme tekniği olarak bilinmektedir. Bu tip öğrenmede veriler etiketlenmez ve veriler bir ön işleme tabi tutulmadan doğrudan algoritmaya uygulanmaktadır. Uygulanan algoritma verileri çözerek yeni etiketlere göre veri kümeleri oluşturmaktadır. Bu tip öğrenmenin kullanılmasındaki neden, veriler hakkında detaylı bilgi edinilmesinin sağlanması amacıyla veri analizinin yapılması veya dağılımının modellenmesinin sağlanmasıdır. Kümeleme (Clustering), Birliktelik (Association), boyutsallığın azaltılması (özellik seçimi) denetimsiz öğrenmenin uygulandığı problemlere örnek olarak verilmektedir. Denetimli öğrenmenin aksine, denetimsiz öğrenmede, kullanılan algoritmaların öğrenimine rehberlik edecek bir veri kümesi bulunmamaktadır (Sammut ve Webb, 2010).

Yarı denetimli ve takviyeli öğrenme yöntemlerinde kullanılan ML algoritmaları, denetimli ve denetimsiz öğrenme algoritmaları kadar yaygın kullanılmamaktadırlar. Şekil 3.11'de denetimli ve denetimsiz öğrenme algoritmalarından güneş tabanlı çalışmalarda sıklıkla kullanılanlar listelenmiştir (Voyant ve ark., 2017). Yarı denetimli öğrenme, denetimli ve denetimsiz öğrenmenin özelliklerinin birleştirildiği bir stratejiyi benimsemektedir. Verileri işleyerek bir ön işlemden geçirmenin ve etiketlemenin bir süreci ve maliyeti bulunmaktadır. Ham ve etiketlenmemiş veriler bu anlamda toplanması ve saklanması oldukça kolay olmaktadır. Yarı denetimli modellerin eğitim sürecinde hem etiketli hem de etiketlenmemiş verileri kullanılmaktadır. Modelleme

sürecinde çoğunlukla etiketlenmemiş veriler kullanıldığından insan önyargısının ve harcanan zamanın azaltılmasına katkıda bulunmaktadır. Takviyeli öğrenmede ise modele girdi olarak hiçbir verinin verilmediği, bunun yerine kullanılan algoritmanın durumu kendi başına çözmesi gerektirdiği durumlarda uygulanmaktadır (Sammut ve Webb, 2010).

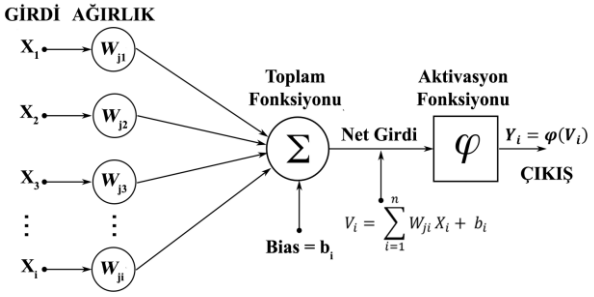


Şekil 3.11. Yaygın kullanılan ML algoritmaları (MathWorks, 2018)

Son yıllarda yapılan güneş tabanlı sistemlerin tahmin çalışmaları hem verilerin analizi hem de kullanılan yöntemlerin özellikleri göz önüne alındığında büyük veri uygulaması olarak görülmektedir. Tahmin çalışmalarında ML'si en fazla tercih edilen yöntem olmakta ve YSA tahmin kesinliği en fazla olan ML algoritmalarının başında gelmektedir (de Freitas Viscondi ve Alves-Souza, 2019). Tez çalışmasında yararlanılan meteorolojik, coğrafik ve kategorik verilerin sayısal olması ve hangi sınıfa ait olduklarının bilinmesi sebebiyle, ML yöntemlerinden denetimli öğrenme içinde yer alan regresyon tabanlı algoritmalardan faydalanılmıştır.

3.3.1. YSA algoritması

Yapılan literatür taramalarında, YSA algoritması güneş tabanlı çalışmalarda kesinliği ve doğruluğu sebebiyle en fazla yararlanılan yöntemlerin başında gelmektedir. Birden fazla ML algoritması kullanılarak yapılan karşılaştırmalı çalışmalarda daha küçük hatalar daha kesin tahmin sonuçlarına ulaşıldığı ve bu algoritmalardan KGI tahmininde en başarılı tahmin modellerinin tartışmasız YSA ile elde dildiği belirtilmiştir (Voyant ve ark., 2017). Bu tez kapsamında Akdeniz Bölgesinin saatlik güneş ışınımın tahmin edilmesinde ana yöntem olarak YSA algoritması kullanılmıştır. YSA, insan beyninin sinir hücresi modellenerek geliştirilen bir ML algoritması olarak tanımlanmaktadır. YSA ile geliştirilen yapay sinir hücresi (nöron) beynin çalışmasına benzer şekilde birden fazla katmanın ve sinir hücresinin birbirlerine bağlanması ile oluşturulan veri tabanlı yapılar olup, insanın öğrenme ve hızlı karar vereme yeteneklerini modelleyerek karmaşık problemlerin çözümünde kullanılması amaçlanmıştır. Tek katmanlı algılayıcının yapısı Şekil 3.12'de verilmiştir. Nöronun tasarımında girişler, ağırlıklar, toplam fonksiyonu, aktivasyon fonksiyonu ve çıkış olmak üzere beş temel yapı bulunmaktadır. Geliştirilen sinir hücresi, insanın biyolojik ve düşünsel yapısından hareketle, benzer bir yapıda çalışmasını sağlayacak şekilde bilgisayar tarafından modellenmiştir. Hesaplama yaparken bilgisayarların kullandığı klasik matematiksel yapıdan farklı çalıştığından doğrusal olmayan ve klasik hesaplama yöntemleri ile hesaplanamayan problemlerin çözümünde oldukça başarılı sonuçlara ulaşılmaktadır. Geleceğe yönelik tahmin, örüntü tanıma, zaman serisi ve veri desenlerinin analizi gibi problemler YSA'nın sıklıkla kullanıldığı çalışma alanları olarak bilinmektedir. Karmaşık veriler ile çalışırken oldukça yüksek performans seviyelerine ulaşılmakta ve esnek yapısı sayesinde farklı ML algoritmaları ile hibrit olarak çalışabilmektedir (Taşdemir ve ark., 2017).



Şekil 3.12. Basit tek katmalı algılayıcının yapısı

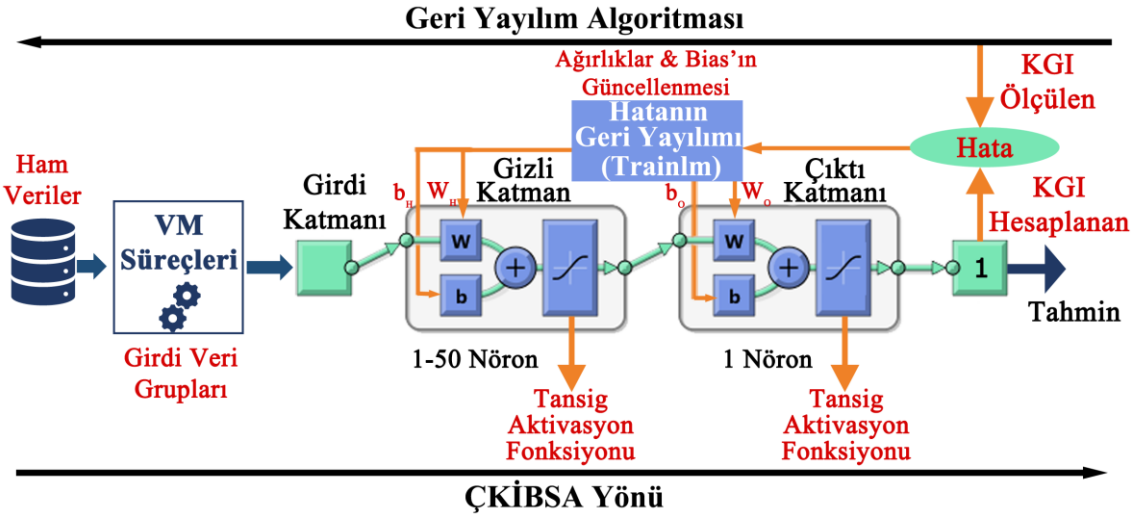
1950’den bu yana farklı amaç ve alanlarda kullanılmak üzere sinir ağı mimarileri geliştirilmiştir. Her mimaride nöronlar arasındaki bağlantının şekli ve öğrenme durumlarına göre farklı isimlerle adlandırılmıştır. Genel olarak YSA, nöronlar arasındaki bağlantıların akış yönüne göre ileri beslemeli (Feed-Forward) ve geri beslemeli (Recurrent) ağlar olmak üzere iki kısma ayrılmaktadırlar. İleri beslemeli ağ mimarisinde giriş katmanından gelen her girdi sinyali belirlenen ağırlık değerleri ile ilgili katmanda hesaplanarak bulunan sonuç bir sonraki katmandaki nöronlara iletilerek ağın tüm sinir hücreleri arasında bir bağlantı yapılmaktadır. Sinir hücreleri arasında gerçekleşen işlemler katmanlar arasında girişten çıkışa tek yönlü olarak ilerlemektedir. Her katmana giriş yapan sinyal ilgili katmanın ağırlık değerleri ile hesaplanarak bir sonraki katmana iletilmektedir. Çıkış katmanına gelindiğinde hesaplanan değerler ile hedeflenen performans değerleri karşılaştırılarak modelin ağ hatası belirlenmektedir. Hata oranına göre her bir katmandaki ağırlıklar tekrardan güncellenmektedir. Tek Katmanlı İleri Beslemeli ve Çok Katmanlı İleri Beslemeli ağ mimarileri bu yapıda inşa edilmişlerdir. Geri beslemeli ağlar ise en az bir geriye dönük bağlantıya sahip olmaları sebebiyle sinyalin döngüler kullanarak hem ileri hem de geri hareket ediyor olmasından dolayı ileri beslemeli ağlardan ayrılmışlardır. Geri beslemeli ağlarda en az bir nöronun çıkışı kendisine veya diğer nöronlara girdi olarak ayarlanmaktadır. Bağlantılar, katmandaki hücrelerde olabileceği gibi katmanlar arasındaki hücrelerde de olabileceğinden doğrusal olmayan dinamik bir davranış göstermektedirler. Katmanlar arasında farklı mimarilere sahip RNN, Hopfield ve Elman ağları gibi geri beslemeli ağlar bulunmaktadır. Bu tip ağlar genellikle denetimsiz öğrenme kurallarının uygulandığı problemlerde kullanılmaktadır (Hagan ve ark., 2002).

Tek katmanlı sinir ağları, günümüz problemlerinin çözümünde yetersiz kalmaktadırlar. Bu sebeple, zor ve karmaşık problemlerin çözümünde çok katmalı sinir ağı mimarileri kullanılmaktadır. Ayrıca öğrenme türüne göre uygun mimari yapının seçilmesi gerekmektedir. Bu sebeple, tez çalışmasında eldeki veri yapısına uygun ve başarı performansı yüksek ileri besleme mimari yapısı dayanan denetimli öğrenme algoritması içinde yer alan ÇKİBSA ML algoritması kullanılmıştır. Geliştirilen YSA modellerinde oldukça fazla veri havuzu ile çalışıldığından ağın kolay yorumlanabilmesi ve modelleme süresini kısaltmak adına bir girdi katmanı, bir gizli katman ve bir çıktı katmanından oluşan, 3 katmanlı bir yapı tercih edilmiştir. YSA tabanlı denetimli öğrenmede sıklıkla geri yayılım (backpropagation), delta, genişletilmiş delta, algılayıcı

gibi öğrenme kurallardan faydalanılmaktadır. Öğrenme kuralı, ağın çıkışında hesaplanan değerlere cevap olarak bellekte var olan ağırlıkların hepsinin veya bazılarının değiştirilmesini sağlayan denklem olarak bilinmektedir. Bu kurallar ile ağın eğitimi gerçekleştirilmekte ve girdi verileri ile hedeflenen çıktıya ne oranda yaklaşıldığı hesaplanmaktadır. Her iterasyonda katman düğümlerinin her biri için ağırlıklar hata sinyalinin durumuna göre yeniden ayarlanmaktadır. ÇKİBSA'nda geri yayılım öğrenme kuralından faydalanılmıştır. Geri yayılımın temelinde, girdi ve çıktı arasındaki doğrusal olmayan ilişkiye ağırlık değerlerini ayarlayarak yaklaşma hedeflenmektedir. Algoritma olarak tasarımı her bir sinir ağı katmanı arasında tam bir bağıntının oluşması şeklinde açıklanmaktadır. Tasarım ve çalışma mantığı açısından diğer öğrenme kurallarına ve sinir ağı modellerine kıyasla özel bir ağ olarak tanımlanmaktadır. Sinir ağı çalıştığında yayılma ve uyum gösterme gibi iki temel prensibi yerine getirerek çıkıştan girişe doğru hataları azaltmak gerekmektedir. Bunun için her bir bağlantı noktasında ağırlıkların yeniden hesaplaması yapılmaktadır. Geri yayılım ağının her katmanında doğrusal olmayan ve paralel olarak işleme sokulan düğüm noktaları bulunmaktadır. Geliştirilen model çalıştırıldığında, ÇKİBSA geri yayılım algoritması ile birlikte statik olmayan davranış sergilemektedir. Bu nedenle birçok alanda uygulama geliştirilebilmektedir (Hagan ve ark., 2002; Şahin, 2019).

ÇKİBSA hatayı minimize edecek şekilde çalıştığından tahmin performansını önemli oranda arttırmıştır. KGI tahmin modellerinde geliştirilen sinir ağının mimari yapısı Şekil 3.13'te verilmiştir. Modellerinin geliştirmesinde Matlab R2017b programından faydalanılarak bir kod yazılmıştır. Sinir ağında 7 il için GR1-G6 arasında girdi veri grupları kullanılırken, sadece bir il için GR1-GR7 arasında veri grupları kullanılmıştır. Her bir grup veri seti için gizli katmanda 1-50 arasında nöron kullanılarak, her bir nöron için 5 tekrar ve her tekrarda maksimum 1000 epok (epoch)'tan oluşan bir eğitim modeli tasarlanmıştır. Geliştirilen ÇKİBSA modellerinde hesaplanan net girdi (N_i) formülü denklem 3.1' de verilmiştir. Bu formülde X_j girdi verisini, W_{ij} ağırlıkları, b_i biası temsil etmektedir. Girdi katmanından çıktı katmanına doğru her bir nöron için bu değerler hesaplanmaktadır.

$$N_i = \sum_{j=1}^{m} (W_{ij} X_j) + b_i \qquad (3.1)$$



Şekil 3.13. Tahmin modellerinde geliştirilen ÇKİBSA mimari yapısı (Guher ve ark., 2020)

Sinir ağında toplam fonksiyonunda net girdi hesaplandıktan sonra bir aktivasyon fonksiyonu kullanılarak etkinleştirmeleri gerekmektedir. Aktivasyon fonksiyonu, toplam fonksiyonun çıkış değerinin değişmesi için kullanılan ve genellikle doğrusal olmayan bir fonksiyondur. Bu sebeple seçilecek aktivasyon fonksiyonunun ağın performansı üzerinde önemli bir etkisi bulunmaktadır. Kullanılan veri setine göre en uygun aktivasyon fonksiyonuna çeşitli denemeler sonucunda karar verilmektedir (Hagan ve ark., 2002). Yaygın kullanılan aktivasyon fonksiyonlarının formülleri denklem 3.2-3.4'te verilmiştir. Yapılan denemeler sonucunda girdi katmanı ile gizli katman ve gizli katman ile çıktı katmanı arasında hiperbolik tanjant sigmoid aktivasyon fonksiyonunun (Tansig) kullanılmasına karar verilmiştir. Kullanılan aktivasyon formülü net girdi hesaplama sonuçlarını -1 ile +1 aralığında ölçeklendirilmiştir. Aktivasyon fonksiyonun seçim sürecinde gizli katmanda lojistik sigmoid (Logsig) veya Tansig fonksiyonu, çıkış katmanında ise Tansig veya doğrusal (Purelin) fonksiyonlarının kullanılabilir olduğu belirlenmiştir. Farklı fonksiyon kombinasyonlarında ağın performansının önemli ölçüde etkilendiği gözlemlenmiştir. En iyi performans her iki katmanda Tansig fonksiyonun kullanımıyla elde edilmiştir.

$$\text{Tanjant Sigmoid}(N_i) = \left(\frac{2}{1+e^{-2N_i}} - 1 \right) = \text{Tansig}(N_i) \qquad (3.2)$$

$$\text{Lojistik Sigmoid}(N_i) = \left(\frac{1}{1+e^{-N_i}} \right) = \text{Logsig}(N_i) \qquad (3.3)$$

$$\text{Doğrusal}(N_i) = N_i = \text{Purelin}(N_i) \quad (3.4)$$

Denetimli öğrenmede YSA kullanılarak ilgili veri setinden bir karara varmak amacıyla geliştirilen modelin eğitim süreci öğrenme olarak adlandırılmaktadır. YSA'nda parametrelerin güncellenmesinde kullanılan geriye yayılım algoritması günümüzde en çok tercih edilen eğitim fonksiyonlarının başında gelmektedir. Ancak farklı amaç ve problemlerde kullanılmak üzere geliştirilmiş birçok farklı tipte geri yayılım eğitim fonksiyonu bulunmaktadır. Herhangi bir problemin çözümünde hangi eğitim fonksiyonunun yüksek performans göstereceğini ve eğitim hızını bilmek güç olmaktadır. Eğitim verilerinin sayısı, ağdaki ağırlıklar ve bias'ların sayısı, ağın performans kriteri, problemin karmaşıklık durumu, problemin türü (örüntü tanıma veya regresyon) gibi sebepler doğrudan öğrenme algoritmasını etkilemektedir (Haykin, 1999). Geliştirilen sinir ağında, Levenberg-Marquardt (Trainlm), Bayes Düzenlileştirme (Trainbr), Gradyan İniş (Traingd), Powell-Beale Yeniden Başlatma Eşlenik Gradyan (Traincgb), Fletcher-Powell Eşlenik Gradyan (Traincgf), Polak-Ribiére Eşlenik Gradyan (Traincgp), Ölçekli Eşlenik Gradyan (Trainscg), Tek Adım Sekant (Tarinoss) ve Değişken Öğrenme Oranlı Geri Yayılım (Traingdx) öğrenme algoritmaları karşılaştırılarak, KGI veri setinde en iyi performans gösteren fonksiyonun belirlenmesi hedeflenmiştir. Yapılan denemelerde Trainlm, Trainbr ve Traincgb fonksiyonları ile elde edilen performanslar birbirlerine oldukça yakın bulunmuştur. Trainlm fonksiyonu diğer iki fonksiyondan az da olsa daha iyi başarım gösterdiğinden çalışmada bu fonksiyonun kullanılmasına karar verilmiştir.

Trainlm eğitim fonksiyonu, maksimum komşuluk amacıyla en küçük kareler yöntemini kullanarak doğrusal olmayan fonksiyonlara sayısal bir çözüm bulmak amacıyla Gauss-Newton ve En-Dik İniş yöntemlerinin en iyi yönlerinin alınması ile geliştirilmiştir. Ancak, Gauss-Newton yönteminde en önemli sorunlardan birisi hessian matrisinin ($H = J^T J$) tersine çevrilebilir olduğundan emin olunamamasıdır. Levenberg-Marquardt algoritmasında optimize edilmiş hessian matrisinin değeri denklem 3.5'te verilmiştir. Denklemde, μ kombinasyon katsayısı, I birim matris ve J ağırlık ve biaslara bağlı olarak ağ hatalarının ilk türevlerinin hesaplanması ile elde edilen Jakobian matrisi olarak tanımlanmaktadır. Jakobian matrisinin birinci dereceden türevi hesaplayan formül denklem 3.6'da verilmiştir. Ağın gradyanı denklem 3.7'de gösterilen formül ile

hesaplanmaktadır. Denklemde e ağ hatası vektörü ve w ağırlık yerine kullanılmaktadır (Yu ve Wilamowski, 2011).

$$H = J^T J + \mu I \tag{3.5}$$

$$J = \frac{\Delta e}{\Delta w} \tag{3.6}$$

$$G = J^T e \tag{3.7}$$

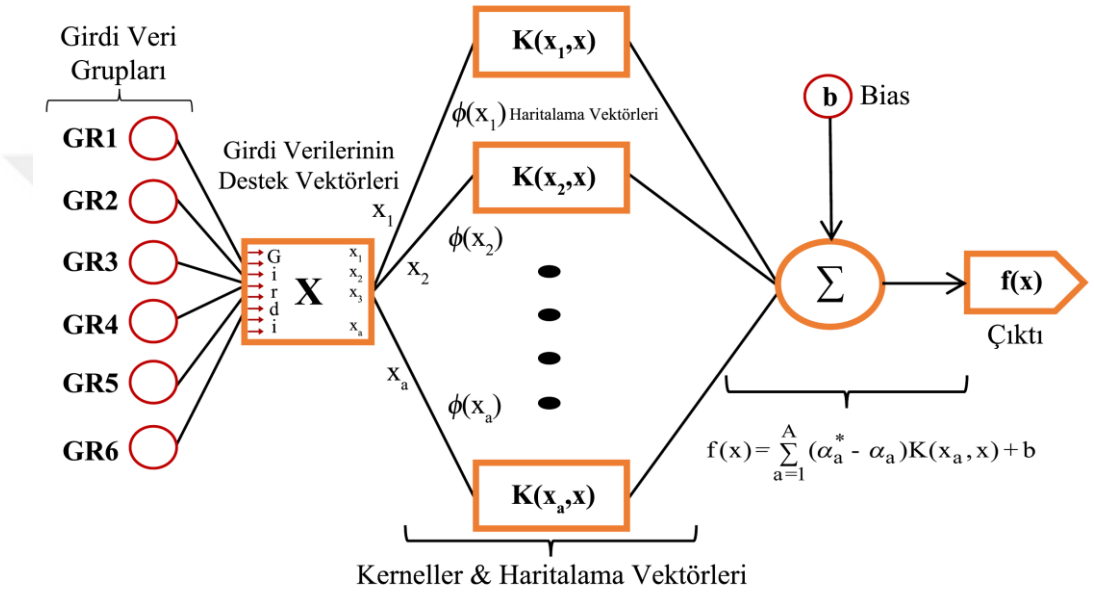
Trainlm algoritmasının formülü denklem 3.8'de verilmiştir. En dik iniş ile Gauss-Newton algoritmasının optimize edilmiş hali olan Levenberg-Marquardt algoritması sinir ağının eğitimi sürecinde iki algoritmanın matematiksel denklemlerinden yararlanarak öğrenme hedefine ulaşmaktadır. μ sıfıra çok yakın olduğunda Gauss-Newton algoritması gibi, μ çok büyük olduğunda en dik iniş yöntemi gibi davranarak ağın ağırlık ve bias değerleri hesaplanmaktadır (Yu ve Wilamowski, 2011).

$$w_{k+1} = w_k - \left[H_k\right]^{-1} G_k \tag{3.8}$$

3.3.2. DVR algoritması

Vapnik tarafından geliştirilen DVM, girdi değişkenleri arasındaki örüntüler hakkında pek fazla bilgi sahibi olunmadığı durumlarda kullanılırken yaygın olarak örüntü tanıma ve sınıflandırma problemleri için tercih edilen bir ML algoritması olarak bilinmektedir. DVM tahmin sürecinde iki sınıf arasındaki en iyi modeli bulmak için girdi verilerinin en küçük alt kümelerine ayrılması ile süreci gerçekleştirmektedir. Yöntem olarak yapısal risk minimizasyonuna dayanan ve denetimli öğrenme stratejisi içerisinde yer alan öğrenme algoritması olarak tanımlanmakta ve istatistik, sinir ağları gibi teknikler bu yöntemde bir arada çalışmaktadır. Aslında DVM, sinir ağlarıyla yakından ilişkilidir ve sigmoid kernel fonksiyonu tabanlı geliştirilen model, çift katmanlı-ileri beslemeli YSA'na benzer bir davranış sergilemektedir. Denetimli öğrenme içerisinde yer alan ML algoritmalarından hataları girdi boyutsallığına göre değil de veri setinin marjına göre ayırmasıyla ayrılmaktadır. Bu nedenle veri setinin çok büyük olduğu durumlarda bile iyi performans göstermektedir (Vapnik, 2000).

Kullanılan yöntemin regresyon tabanlı çok sınıflı tahmin problemlerine uyarlanmasında yaşanan zorluklardan, DVM tabanlı DVR algoritması geliştirilmiştir. DVR'nda amaç ikiden fazla değişkenin olduğu veri setlerinde regresyon problemlerinin çözümüne benzer bir tekniğin kullanılabilmesini sağlamaktır. Bu sayede çok boyutlu özellik kümelerinden oluşan veri setlerinin doğrusal regresyon fonksiyonunun hesaplanması gerçekleştirilebilecektir (Akıl, 2019). Geliştirilen DVR tabanlı tahmin modelinin mimari yapısı Şekil 3.15'te verilmiştir.



Şekil 3.14. DVR tabanlı geliştirilen modelin mimari yapısı (Guher ve ark., 2020)

Geliştirilen modellerde girdi verileri arasındaki boşluklar mümkün olduğunca geniş tutulup en ideal konumları bulunarak hataların minimize edilmesine çalışılmaktadır. Modelleme çalışmasında kullanılan veri seti $\{(x_a, y_a),\ a = 1,2,3,...,M\}$ denklemi ile tanımlanmaktadır. Bu denklemde, $x_a \epsilon R^d$ girdi vektörünü, $y_a \epsilon R$ x_i karşılık gelen çıktı vektörünü, M toplam eleman sayısını temsil etmektedir (Ko ve Lee, 2013). DVR tabanlı modelin doğrusal fonksiyonunun denklemi 3.9'da verilmiştir. Denklemde, W ağırlık vektörü, b ise bias ve $\varphi(x)$ ise çok boyutlu değişkenleri iki boyutlu grafiğe dönüştüren haritalama fonksiyonu yerine kullanılmıştır.

$$y = f(x) = W^t \varphi(x) + b \qquad (3.9)$$

DVR modellerinde kullanılan hata fonksiyonu formülü denklem 3.10'da gösterilmiştir. Modelleme sürecinde C sabiti ve ε - epsilon kullanıcı tarafından belirlenen parametrelerdir ve eğitim sürecinde ayarlanarak tahmin kesinliğine önemli katkılar sağlamaktadırlar.

$$R(x) = \frac{1}{2} \| W \|^2 + C \frac{1}{N} \sum_{i=1}^{N} | y_i - f(x_i) |_{\varepsilon} \quad (3.10)$$

Modellerde hata fonksiyonunu en aza indiren formül denklem 3.11'de verilmiştir. Denklemde geçen α_a^* ve α_a lagrange çarpanları olarak adlandırılmaktadır ve eğer sıfırdan farklı bir değere sahipse bunlara destek vektörleri denilmektedir. Lagrange çarpanları DVR tabanlı geliştirilen modeller için hayati öneme sahip değerler olarak bilinmektedir. Denklemde, $K(x_a,x)$ formülü kernel (çekirdek) fonksiyonudur ve görevi girdi olarak alınan bilgileri kullanılabilir bir forma dönüştürmektir. Geliştirilen modellerde farklı tiplerde kernel fonksiyonları kullanılmaktadır. KGI tahmin modelinde Polinomial (POL), Normalize Polinomial (NOR-P) ve Radial Basis Functions (RBF) olmak üzere 3 farklı tipte kernel kullanılmış ve formülleri sırasıyla denklem 3.12-3.14'te verilmiştir.

$$f(x) = \sum_{a=1}^{A} (\alpha_a^* - \alpha_a) K(x_a, x) + b \quad (3.11)$$

$$K(x_a, x) = [(x_a x) + 1]^d \quad (3.12)$$

$$K(x_a, x) = \frac{[(x_a x) + 1]^d}{\sqrt{(x_a)^2 (x)^2}} \quad (3.13)$$

$$K(x_a, x) = \exp\left(-\frac{1}{2\sigma^2} \| x_a - x \|^2\right) \quad (3.14)$$

3.3.2.1 LibSVM kütüphanesi

Standart DVM algoritması olasılıksız ve ikili doğrusal bir sınıflandırıcı olarak tahmin modellerinde kullanılmaktadır. Kernel fonksiyonları yardımıyla doğrusal olmayan sınıflandırma yapabilmekte ve sınıfları doğrusal olarak ayırmak için yüksek

boyutsal özellik kazandırılabilmektedir. LibSVM kütüphanesi içinde var olan birçok kernel fonksiyonları sayesinde çok katmanlı sınıflandırma ve farklı çapraz doğrulama gibi işlemleri kolaylıkla yerine getirebilmektedir. Bu sayede DVM'nın özelliklerini ayrıntılı bilmeyen araştırıcılara kolay bir şekilde uygulama geliştirebileceklerini sağlamak amacıyla oluşturulan kütüphane yazılımıdır ve çeşitli uygulama yazılımlarına kolaylıkla entegre edilebilmektedir (Chang ve Lin, 2011).

LibSVM yaygın olarak tek sınıflı DVM problemlerinde, iki ya da çok sınıflı DVM ve DVR problemlerinde son yıllarda çok sık kullanılmaktadır. Ancak, KGI tahmin modellerinde sık kullanılmayan bir yöntem olduğu için bu tez çalışmasında tercih edilmiştir. DVR tabanlı bir başka yöntem olan LibSVM kütüphanesi kullanılarak tahmin modelleri geliştirilmiş ve klasik DVR algoritması ile karşılaştırılmıştır. LibSVM ile geliştirilen tahmin modellerinde, Epsilon DVR (E-DVR) ve Nu-DVR olmak üzere iki farklı DVR tipi kullanılmış ve kernel fonksiyonu olarak RBF faydalanılmıştır. Nu-DVR'nin Epsilon DVR'den farkı epsilon hata parametresi yerine tahmin kesinliğini arttırmada nu parametresinin kullanılarak modelin daha kolay yorumlanabilmesini sağlamaktır.

3.3.3. K-NN algoritması

K-NN algoritması, seçilen veri setinde daha önceki verilerden K tanesinin komşuluğuna bakılması amacıyla kullanılmaktadır. Algoritma ile K parametresinin aldığı değere bağlı olarak tahminler yapılmaktadır. Modelleme sürecinde eğer K değeri bir sayısına eşit ise sadece en yakın komşunun bulunduğu tahmin sınıfına atanmaktadır. K parametresinin değeri arttıkça veri setinde yer alan tüm veriler dikkate alınmakta ve oylamaya yöntemine göre verilerin seçimi yapılmaktadır. Oylama sürecinde her bir komşu değeri için ağırlıklandırma yapılarak tahmin sınıflandırması yapılan örneğe daha yakın olan komşu örneklerin bulunması sağlanmaktadır. Modele girilen yeni örnek verisi, eğitim verisindeki örneklerin benzerliğe göre sınıflandırılmaktadır. Sınıflandırma problemlerinde çoğunlukla kullanılan bu algoritmanın regresyon temelli bir uygulaması çalışmada tercih edilmiştir. K-NN non-parametric tembel bir ML algoritması olarak bilinmektedir ve denklem 3.15'te kullanılan algoritmanın formülü verilmiştir (Hastie ve ark., 2009). Denklemde her $N_k(x)$'te x verisinin komşusu K olarak alınmıştır ve y_i değeri her x_i girdi verisi için çıktıyı temsil etmektedir.

$$f_k(x)=\frac{1}{K}\sum_{i\in N_k(x)} y_i \tag{3.15}$$

K-NN algoritması kolay uygulanabilir olması, bozuk ve gürültülü verilerden daha az etkilenmesi ve yeni bilgilere uyarlanabilir olması gibi nedenlerle birçok alanda modeller geliştirilmiştir. Geliştirilen uygulamalarda asıl amaç tahmin yapılması istenilen her yeni verinin daha önceki veriden K tanesinin komşuluğunu ayarlamaktır. Seçilen veri seti ile hedef veri setindeki değerler arasındaki uzaklık mesafesi hesaplanmaktadır. Bu süreçte K değerinin modelin performansında çok önemli etkisi bulunmaktadır. Bazı tahmin modellerinde K parametresinin artırılması performansa olumlu etkisi olurken, bazılarında ise tam tersi yönde bir etki gösterdiği gözlemlenmiştir. Bu sebeple K değerinin seçiminde dikkatli olunması gereklidir ve çok yüksek değerler kullanılır ise model aşırı uyma eğiliminde olmaktadır. Komşu değerler arasındaki mesafe hesaplanırken Öklid, Minkowski, Manhattan, Chebyschev gibi uzaklık fonksiyonları kullanılmaktadır (Hastie ve ark., 2009). Geliştirilen modelde farklı uzaklık fonksiyonları denenenmiş, tahmin başarısına etkisi sebebiyle Öklid fonksiyonunun kullanılmasına karar verilmiştir. Öklid, K-NN ile yapılan modelleme çalışmalarında en fazla yararlanılan uzaklık fonksiyonu olarak bilinmektedir. İki nokta arasındaki doğrusal uzaklık olarak tanımlanan Öklid fonksiyonunun formülü denklem 3.16'da verilmiştir. Tahmin modellerinde K değerinin 1, 2, 3, 4, 6 ve 10 olarak alınıp kullanılmasına karar verilmiştir. 10'un üzerindeki değerlerde tahmin modeli performanslarında herhangi bir artış gözlenmemiş, tahmin sonuçları aşırı uyuma girmiştir.

$$\text{Öklid}=\sqrt{\sum_{i=1}^{K}(x_i-y_i)^2} \tag{3.16}$$

3.3.4. M5 kuralları algoritması

ML algoritmaları içerisinde ağaç tabanlı öğrenme algoritmaları, en çok kullanılan denetimli öğrenme algoritmaları olarak bilinmektedirler. Sınıflandırma ve regresyon problemlerinin birçoğu için çözüm geliştirilebilmektedir. Bu algoritmalar içerisinde en çok kullanılan karar ağaçları algoritmasında, veri seti içerisindeki örnek

veriler basit karar adımları ile küçük gruplara bölünmekte ve her bölme işlemi sonucunda birbirine benzeyen veriler gruplanmaktadır. Süreç içinde basit karar verme yöntemleri uygulanarak büyük miktarlardaki verilerden çok küçük veri grupları oluşturarak en iyi tahmin modelinin geliştirilmesi sağlanmaktadır. Düğümlerin oluşturulması karar ağaçlarının temel işlevi olarak bilinmektedir. Bu süreçte ağacın dengeli bir şekilde dallanması ve en iyi niteliğin düğüm olarak seçilmesi gerekmektedir. Karar ağacı oluşturmak için geliştirilen ID3, C4.5, MARS, Sınıflandırma ve Regresyon Ağaçları (CART) gibi algoritmalar amaca uygun kullanılabilmektedir (Witten ve Frank, 2005).

Sınıflama ve regresyon problemlerinin çözümünde kullanılan CART algoritması, kategorik ve sayısal veri setlerindeki değişkenlere uygulanarak karar ağaçları temelli non-parametrik bir yöntem olarak bilinmektedir. Üzerinde çalışılan veri setindeki bağımlı değişkenler kategorik ise Sınıflandırma Ağaçları, sayısal ise Regresyon Ağaçları olarak adlandırılmaktadır. Quinlan tarafından geliştirilen M5 algoritması ise CART'nın gelişmiş versiyonu olarak bilinmektedir (Quinlan, 1992). Ağaçlardan kural üretmeyi amaçlayan yaygın kullanılan bir ML algoritmasıdır. Geliştirilen modelde her yaprak için çok değişkenli doğrusal bir yapı kullanılarak regresyon ağaçları oluşturulmakta ve ağacın düğümleri arasındaki çıktı parametrelerinde arzulanan hata miktarını en düşük seviyede hesaplanmasını sağlayan özellikler seçilerek amaca ulaşılmaktadır. Ağaç yapraklarının bağımlı ve bağımsız değişkenleri arasında ilişki geliştirilmekte ve her yaprakta ikili bir karar ağacı yapısına dayanan doğrusal regresyon modeli oluşturulmaktadır. Algoritmanın çalışma mantığı gereği karar ağaçları ve doğrusal regresyon gibi iki temel yapı üzerine inşa edildiği görülmektedir. Ağacın yaprakları arasında en iyi yaprağın kuralı hesaplanırken budama ve bölme gibi iki aşamalı bir süreçten geçmektedir. Bölme sürecine, mevcut veri seti alt kümelerine ayrılmaktadır. Bu süreçte ayrıca, yaprak düğümlerinde doğrusal regresyon algoritmasından faydalanılarak sayısal özelliklerin sürekli tahmin edilmesi amaçlanmaktadır (Wang ve ark., 2017).

Karar ağacı tabanlı modellerde ilk kök düğümün bulunduğu en tepeden aşağıya doğru bir ağaç yapısı oluşturularak girdi verilerinden benzer değerler içeren alt kümelere bölünmesi sağlanmaktadır. Verilerin benzerliklerinin hesaplanmasında yani ağaç oluşturmada standart sapma kullanılmaktadır. Denklem 3.17'de standart sapmanın formülü verilmiştir. İlgili formülde, X girdi verisini, $\overline{X}$ girdi verilerinin ortalamasını, S

standart sapmayı temsil etmektedir. Homojen dağılıma sahip bir veri setinin standart sapması 0 olarak bulunmaktadır. Ağaçta en homojen dalı bulmak, hesaplanan en yüksek Standart Sapma Azaltma (SDR) sonucuna sahip değişken verisiyle ilişkilendirilmiştir. Bir düğüme ulaşan örneklerin değerlerinde çok az bir değişme olursa ya da örnek sayısı çok azalırsa bölme süreci sonlandırılmaktadır. Standart sapma azaltma (SDR) formülü denklem 3.18'de verilmiştir. Denklemde, T düğüme ulaşan özellik değerleri kümesini, T_i bölünmüş düğümden alınan değişken değerleri olarak adlandırılmaktadır (Witten ve Frank, 2005).

$$S = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(X_i - \overline{X})} \tag{3.17}$$

$$SDR = S(T) - \sum_{i}\frac{|T_i|}{|T|}S(T_i) \tag{3.18}$$

Tez çalışmada, M5 algoritmasının kural-tabanlı geliştirilen bir modeli kullanılmıştır. M5 kuralları olarak adlandırılan bu yöntemde, en iyi yaprağın (kural) saklandığı bir dizi M5 model ağacı oluşturulmakta ve her tekrarda en iyi kuralın sahip olduğu veri seti bir sonraki ağaç oluşturmadan eğitim veri setinden kaldırılması sağlanmaktadır. M5 algoritması yapısı gereği tek bir karar ağacı oluştururken, M5 kuralları algoritması ile her döngüde tam bir ağaç oluşturulması sağlanmaktadır (Witten ve Frank, 2005).

3.4. İstatistiksel Performans Değerlendirme Metrikleri

KGI'nı tahmin etmek amacıyla geliştirilen ML algoritmaları modellerinin hem kendi hem de birbirleri arasındaki tahmin performanslarını değerlendirmek maksadıyla, güneş tabanlı yayınlarda sıklıkla kullanılan istatistiksel hata ölçüm ve analiz metriklerinden yararlanılmıştır. Ortalama Kare Hata (MSE), Kök Ortalama Kare Hata (RMSE), Ortalama Mutlak Hata (MAE) ve Simetrik Ortalama Mutlak Yüzde Hata (SMAPE) çalışmada kullanılan hata ölçüm istatistikleridir. Korelasyon Katsayısı (R) ve Belirleme Katsayısı (R^2) gibi iki farklı istatistiksel analiz yönteminden faydalanılmıştır. MSE, herhangi bir ML algoritması kullanılarak geliştirilen modelin tahmin performansını pozitif olarak ölçer ve sıfıra yakın olan modelin daha iyi bir performans

gösterdiği söylenebilmektedir. Denklem 3.19'da MSE değeri hesaplanabilmektedir (Guher ve ark., 2020).

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(O_i - P_i)^2 \tag{3.19}$$

RMSE, tahmin değerlerinin ölçüm değerlerinden ne kadar saptığının bir ölçüsü olarak bilinmektedir. Tahmin hatalarının standart sapması olan RMSE'nin MSE'ye benzer şekilde sıfıra yakın değerleri iyi performansın ölçütüdür ve denklem 3.20'de formülü verilmiştir (Şahin, 2019).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(O_i - P_i)^2} \tag{3.20}$$

MAE, ölçülen ve tahmin edilen değerler arasındaki farkın ölçüsü olarak bilinmekte ve yönleri dikkate almadan tahmindeki hataların ortalama büyüklüğünü ölçen, regresyon ve zaman serisi problemlerinde sıklıkla kullanılan bir istatistiksel ölçektir. Denklem 3.21'de verilen MAE'ye göre sıfıra yaklaşan modelin diğer modellere göre daha iyi performans gösterdiği ifade edilebilmektedir (Premalatha ve Valan Arasu, 2016).

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|O_i - P_i| \tag{3.21}$$

Regresyon ve zaman serileri modellerinde tahminlerin doğruluğunu ölçmek amacıyla yüzde hataları olarak bilinen ölçekler sıklıkla kullanılmaktadır. MAPE yüzde hatası, ölçekten bağımsız bir tahmin hatası ölçütü olarak ölçülen ve tahmin edilen değerler sıfır olduğunda veya sıfıra oldukça yakın bir değer aldıklarında hatalı sonuçlar vermektedir. MAPE sonuçları değerlendirilirken %10 altındaki tahminler mükemmel kabul edilirken, %10-%20 arası iyi, %20-%50 arası kabul edilebilir ve %50 üzeri ise kötü olarak ifade edilmektedir. Tez çalışmasında hesaplanan ve ölçülen değerlerde sıfır ve sıfıra yakın değerler bulunduğundan bu sorunun üstesinden gelmek için SMAPE

yüzde ölçeği kullanılmıştır. Kullanılan ölçeğin formülü denklem 3.22'de verilmiştir (Guher ve ark., 2020).

$$SMAPE = \frac{100}{n} \sum_{i=1}^{n} \frac{|O_i - P_i|}{(|O_i| + |P_i|) * 0.5} \tag{3.22}$$

R, ölçülen ve tahmin edilen değerler arasında doğrusal bir ilişki olup olmadığını gösterir ve -1 ile 1 arasında bir değer almaktadır. Negatif değerler negatif yönde ilişkiyi gösterirken, pozitif değerler ise pozitif yönde ilişki durumunu ifade etmektedir. Çıkan sonuç, 1 veya -1 yaklaştığında mükemmel derecede bir ilişkiyi, 0'a yaklaştıkça ilişki durumunun olmadığı ya da çok zayıf bir ilişkinin olduğu anlamı çıkarılmaktadır. Denklem 3.23'de R'nin formülü verilmiştir (Taşdemir ve ark., 2017).

$$R = \frac{\sum_{i=1}^{n} (O_i - \overline{O})(P_i - \overline{P})}{\sqrt{\sum_{i=1}^{n} (O_i - \overline{O})^2} \sqrt{\sum_{i=1}^{n} (P_i - \overline{P})^2}} \tag{3.23}$$

R^2, korelasyon katsayısına benzer şekilde ölçülen ve tahmin edilen değerler arasındaki doğrusal ilişkiyi belirlemede yaygın olarak kullanılan istatistiksel analiz yöntemlerinin başında gelmektedir. Bu yöntemde hesaplama sonucu 0 ile 1 arasında pozitif değerler oluşmakta ve çıkan sonuç 1'e yaklaştıkça tahmin ile ölçülen değerler arasındaki bağımlılığın kuvvetli olduğu anlamına gelmektedir. Denklem 3.24'de hesaplamada kullanılan formül gösterilmiştir (Taşdemir ve ark., 2017).

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (O_i - P_i)^2}{\sum_{i=1}^{n} (O_i - \overline{O})^2} \tag{3.24}$$

Yukarıda verilen denklemlerde, O_i, P_i, $\overline{O}$ ve $\overline{P}$ sırasıyla ölçülen, tahmin edilen, ölçüm ve tahmin ortalama değerlerini temsil etmektedir.

4. VERİ MADENCİLİĞİ SÜREÇLERİ

Bu bölümde, Akdeniz bölgesindeki sekiz ilin belirlenen meteorolojik ölçüm istasyonlarından alınan verilerinin modelleme çalışmasına girmeden önce yapılan veri madenciliği ön işlem prosesleri ayrıntılı olarak anlatılmıştır. Süreç sonunda geliştirilen tahmin modellerinde kullanılacak formatta meteorolojik, coğrafik ve kategorik girdi verileri belirlenmiştir. Veri madenciliği süreçlerinden yararlanılarak veri setinden optimal girdi değişkenlerinin belirlenmesi sağlanmış, birden fazla yöntem, teknik ve fonksiyonlardan faydalanılarak girdi özelliklerinin farklı şekillerde gruplanması süreçlerine değinilmiştir.

4.1. Tahmin Modellerinde Kullanılacak Verilerin Ön İşlem Süreçleri

Günümüzde veri boyutu ve verilerin özellik sayısı oldukça büyüdüğünden ML tabanlı tahminsel modelleme çalışmalarında doğrudan kullanılması çıktı performansını negatif yönde etkilediği bilinen bir gerçekliktir. Daha doğru bir ifade ile ML algoritmalarının veri tabanlı yapılar olduğu bilindiğinden veri setinin hiçbir işlemden geçirilmeden kullanılması tahmin başarısını önemli oranlarda düşürdüğü anlaşılmıştır. Bu nedenle toplanan veri setinin geliştirilen tahmin modellerinde kullanılacak şekilde hazırlanması gerekmektedir. Genellikle denetimli öğrenmede kullanılan veri ön işleme adımları, veri madenciliği süreçlerinin en temel yapıları olarak bilinmektedir. Veri ön işleme, ML algoritmaları ile tahmin modelleri geliştirilmeden önce veri setinin amaca uygun hazırlanması ve düzenlenmesi, eksik verilerin belirlenmesi ve tamamlama işlemleri, tekrar edilen verilerin kaldırılması gibi birtakım veri tabanlı işlemleri içeren süreçlerdir. Bu süreçler aynı zamanda ham verilerin ML tabanlı modellere hazırlık aşamaları olarak bilinmekte ve veri temizleme, dönüştürme, indirgeme ve birleştirme adımlarından oluşmaktadır. Birçok çalışmada önemsenmeyen veya kısmen önemsenen bu süreçler son yıllarda ML modellerinin ve veri madenciliğinin en önemli aşamaları olarak üzerinde en fazla durulan yöntem ve tekniklerden oluşmakta ve modelleme çalışmalarının oldukça büyük bir zaman dilimini kapsamaktadır (Gibert ve ark., 2016).

ML algoritmaları ile geliştirilecek tahmin modellerinin doğruluğunu arttırmada kullanılacak girdi verilerinin kalitesini arttıracak birtakım işlemlerin yapılması genellikle önemsenmeyen, ancak son yıllarda modelleme çalışmalarının en önemli aşaması sayılarak üzerinde en fazla durulan yöntemlerin başında gelmektedir. KGI

tahmin çalışmalarında pek tercih edilmeyen veri madenciliği süreçlerinin temelinde kaliteli veri, kaliteli tahmin sonuçları doğurur fikri yatmaktadır. Kullanılan veri seti bozuk, düzenlenmemiş veya kısmen düzenlenmiş ise ne kadar iyi tahmin modelleri geliştirirsek geliştirelim yüksek düzeyde bir performans seviyesine ulaşmak imkansızlaşmaktadır. Bu tipte bir veri seti ile çalışıldığında bir performans seviyesine ulaşılır, fakat "Daha iyiye ulaşabilir miyim?" sorusu kafalarda kalmaktadır (Witten ve Frank, 2005).

ML yöntemleriyle tahmin yapılırken hedef çıktıya göre alakalı verilerin seçimi ve düzenlenmesi hayati bir süreçtir. KGI'na doğrudan ve dolaylı yoldan etki eden veya edebilecek birçok coğrafik, meteorolojik ve kategorik değişkenler bulunduğundan veri havuzu oldukça geniştir. Geliştirilen tahmin modellerinin verinin türü, yapısı, uzunluğu, büyüklüğü gibi özelliklerden etkilendiğinden hareketle çalışmaya başlamadan ham veri seti üzerinde bir takım ön işlemlerin yapılması amaçlanmıştır. Güneş ışınımı tahmin çalışmalarında nadir kullanılan veri madenciliği yöntemlerinden yararlanılarak girdi verinin hazırlık süreçlerinin düzenlenmesi planlanmıştır. Ayrıca, saatlik verilerle yapılan çalışmaların azlığı ve güneş enerjisi santrallerinin kapsamlı ve detaylı tasarımlarının yapılması planlandığından, Akdeniz bölgesi illerinin MGM'den alınan meteorolojik verilerinin saatlik ölçekte olması kararlaştırılmıştır. Buradan hareketle, yapılan çalışmada her bir il için 2002-2006 yılları arasında ölçülen S, B, N, RH, GS ve KGI meteorolojik verileri talep edilmiştir. Çizelge 4.1'de talep edilen meteorolojik ham verilerin karakteristik özellikleri verilmiştir. Çizelgeden de görüleceği üzere S, B, N ve RH verileri günün her saati ölçülen ortalama verilerden oluşurken, GS ve KGI verileri ise gün doğumu ve gün batımı temel alınarak belirli saat dilimlerinde ölçülmüştür. Ancak, KGI ve GS verilerinin, her bir ilin belirlenen ölçüm istasyonlarından alınan değerlerinin ölçüm zaman aralıklarının başlangıç ve bitiş değerlerinin farklılığı dikkati çekmiş ve yıllık ölçülen saatlik veri sayılarının düzensiz olduğu gözlemlenmiştir. Her iki veri değişkenine ait ölçüm değerlerinin aynı istasyonun ölçümün yapıldığı yıl ve ayında bile değişiklik gösterdiği görülmüştür. Bu sorunun üstesinden gelebilmek için, Akdeniz bölgesinde belirlenen sekiz ilin KGI ve GS verilerini ortak bir ölçüm zaman aralığında ayarlanması planlanmıştır, Ocak ve Şubat ayları 06:00-17:00 saatleri arası, Mart, Nisan, Ekim, Kasım, Aralık ayları 06:00-18:00 saatleri arası ve Mayıs-Eylül ayları arasında 07:00-19:00 saatleri arasındaki verilerin seçilmesi kararlaştırılmıştır. Oluşturulacak veri setinde saatlerin belirlenen aralıklarda seçiminde ölçüm zaman

aralıkları farklılıkları, her bir ayın kaydedilen veri saatlerinin değişkenliği ve kış ve yaz saati uygulamaları etkili olmuştur.

Çizelge 4.1. Akdeniz Bölgesi illerinin ham meteorolojik verilerinin karakteristik özellikleri

Meteorolojik Veriler	Ölçüm Birimleri	Ölçüm Zaman Aralığı	Yıllık Ölçülen Veri Sayısı	En Düşük	En Yüksek	Ortalama
S	°C	24 Saat	8,760	-13.8	43.9	19.2
B	hPa	24 Saat	8,760	876	1,031.8	974.5
N	%	24 Saat	8,760	0	100	52.4
RH	m/s	24 Saat	8,760	0	9.9	2.2
GS	Saat	4:00-20:00	-	0	1	0.6
KGI	Cal/cm^2	4:00-20:00	-	0	90.6	28.6

Tez çalışmasında, veri madenciliği süreçlerinde yararlanılan veri ön işleme adımları temel alınarak girdi verileri düzenlenmiştir. İlk olarak il bazında toplanan meteorolojik verilerin bir Excel dosyasında birleştirilmesi amaçlanmıştır. MGM'den talep edilen verilerin her biri hedef istasyonun farklı ölçüm sensörleri tarafından ölçüldüğünden ayrı dosyalar halinde gönderilmiştir. Veri birleştirilmesinin amacı, farklı veri kaynaklarından elde edilen verilerin birlikte değerlendirilebilmesi için tek bir kaynak üzerinde bir araya getirilmesidir. Matlab programında yazılan bir kod ile her bir ilin ölçüm istasyonlarından talep edilen verilerin tamamı yıl bazında tek bir Excel dosyasında birleştirilmiştir. Bu dosyaya, ölçümün yapıldığı yıl, yılın ayı, ayın günü ve günün saati olmak üzere dört adet kategorik veri dahil edilmiştir. Her bir il için ayrı tahmin modelleri geliştirileceğinden enlem, boylam, yükseklik gibi coğrafik veriler çalışmaya dahil edilmemiştir. 2002-2006 arasındaki her bir yıl için S, B, N, RH verileri gün boyu ölçüldüğünden ortalama 8,760 veri aktarılırken, KGI ve GS verilerinde toplam aktarım sayıları değişmektedir. Süreç sonunda Akdeniz bölgesinin sekiz ilinin ham verileri ayrı ayrı tek bir kaynakta birleştirilmiştir.

Birleştirme sürecinden sonra gürültülü verilerin düzenlenmesi gerekmektedir. Veri temizleme, veri ön işleme süreçlerinin en önemli adımlarından birisi olarak tanımlanmaktadır. Çünkü, eksik, hatalı ve tutarsız ölçülmüş veriler ML algoritmalarının performansını olumsuz yönde etkilemektedir. Eksik ve hatalı ölçülen verileri çıkarmak bir yöntemdir ancak, veri yapısının bozulmasına, sapmalara hatta önemli verilerin veri havuzundan silinmesine yol açmaktadır. Bu nedenle eksik verilerin istatistiksel (ortalama, medyan, mod gibi) veya tahmine dayalı (K-NN, DVM, GA gibi) yöntemlerle tamamlanması en fazla tercih edilen yöntemlerin başında gelmektedir (Famili ve ark.,

1997). İlk olarak veri setinde tamamen hatalı ölçülmüş veriler belirlenerek temizlenmiştir. Ardından Excel'de yazılan bir program kodu ile eksik ve kısmen hatalı ölçülmüş veriler bir önceki ve bir sonraki yılın aynı zaman dilimindeki değerlerinin aritmetik ortalaması alınarak tamamlanmıştır. Program kodu ilk olarak ilgili veri sütunundaki eksik veri satırını bulmakta, ardından bir önceki ve bir sonraki yılın aynı satırındaki değerleri bularak bunların ortalamasını almaktadır. Bu işlemi yaparken önceki ve sonraki yılın ilgili satırında değer kontrolü yapılmakta, yoksa otomatik olarak bir başka yılın ilgili satırındaki dolu hücreye konumlanmaktadır. Ayrıca, düzenleme yapılan satır numaraları kaydedilmekte ve program böyle bir satıra geldiğinde bu satır tekrardan işleme alınmamaktadır. Bu yöntemle tüm meteorolojik değişkenlerin eksik verileri güncellenmiştir. Hesaplanan veriler, Akdeniz bölgesi için talep edilen verilerin yaklaşık %5'i kadardır. Ayrıca bu adımda, KGI ve GS verilerinin ölçüm zaman aralıklarının dışında kalan diğer veriler tablodan çıkarılmıştır. Süreç sonunda her bir yıl için ortalama 4,686 adet temiz, kaliteli ve gürültüsüz veri elde edilmiştir. Her bir il için ML algoritmalarında kullanılmak üzere toplamda 23,442 adet veri düzenlenmiştir.

ML algoritmalarının birçoğu veri dağılımlarından etkilenmektedir. Özellikle regresyon tabanlı algoritmalar dağılımın yönünden oldukça fazla etkilenmekte olup genellikle normal dağılıma sahip veri setlerinde başarılı olmaktadır. Bu nedenle eldeki verilerin tahmin modellerinde kullanılacak şekilde yeniden ayarlanması gerekmektedir. Veri dönüştürme olarak bilinen bu adım, birbirlerinden farklı ölçeklerde kaydedilen verilerin normalizasyon, ölçeklendirme gibi tekniklerle dağılımının düzenlenmesi işlemleridir. Meteorolojik veriler düzensiz veriler olduğundan ve farklı ölçüm aralıklarına sahip olduklarından hiçbir dönüştürme yapmadan modelleme çalışmalarında kullanılması tahmin performansını önemli oranda düşürmektedir. Veri dönüştürme sürecinde en sık yararlanılan yöntem normalizasyon işlemidir. Bu nedenle uygun bir normalizasyon tekniğinden faydalanılarak çalışmalarda kullanılacak verilerin dönüştürülmesi amaçlanmıştır. Ancak literatürde oldukça fazla normalizasyon yöntemi olduğundan tahmin modeli üzerinde en fazla etkiye sahip olan normalizasyon tekniğinin belirlenmesi kararlaştırılmıştır. (Tasdemir ve ark., 2018) yaptıkları bir çalışmada, YSA ML algoritması ile rüzgâr gücünü tahmin ederken yaygın olarak kullanılan dört farklı normalizasyon tekniğinin tahmin performansı üzerindeki etkilerini araştırmışlardır. Rüzgâr tabanlı veri setine sigmoid normalizasyonu uygulanarak geliştirilen modelin daha iyi performans seviyesine ulaştığını göstermişlerdir.

Birden fazla ilin düzenlenmiş girdi verilerine Min-Max, Sigmoid ve Z-Skor normalizasyonu uygulayarak YSA algoritması ile tahminler yapılmıştır. Denklem 4.1-4.3'te sırasıyla çalışmada faydalanılan Min-Max, Sigmoid ve Z-Skor normalizasyon yöntemlerinin formülleri verilmiştir. Formüllerde geçen Y_i ve X_i değerleri sırasıyla normalize edilen ve ham girdi verilerini temsil etmekte, e^{x_i} değeri ilgili verinin doğal logaritmasını hesaplamaktadır. μ_i ve σ_i değerleri ise sırasıyla veri setinin ortalamasını ve standart sapmasını hesaplamak için kullanılmıştır. Min-Max normalizasyonunda veri setinin minimum ve maksimum değerleri bulunarak doğrusal olarak 0-1 aralığında normalize edilmektedir. Sigmoid normalizasyonu, geliştirilen modelin daha önceden görmediği veri setlerinin dağılımından etkilenmeyen bir yöntem olarak bilinmektedir. Hesaplama sonucunda verilerin dağılımı 0 ile 1 veya -1 ile 1 aralığında olmaktadır. Z-Skor normalizasyonunda ise, veri setinin ortalaması ve standart sapması hesaplanmaktadır. Yaygın olarak veri setinin değerlerinin değişmediği sabit durumlarda kullanılmaktadır.

$$Y_i = \frac{X_i - X_{min}}{X_{max} - X_{min}} \tag{4.1}$$

$$Y_i = \frac{e^{x_i} - e^{-x_i}}{e^{x_i} + e^{-x_i}} \tag{4.2}$$

$$Y_i = \frac{(X_i - \mu_i)}{\sigma_i} \tag{4.3}$$

Min-Max normalizasyon yöntemi ile geliştirilen modellerin tahmin doğruluğu diğer normalizasyon teknikleriyle geliştirilen modellerden üstün bulunmuştur. ML algoritmalarında kullanılmak üzere Akdeniz bölgesi veri setinin Min-Max normalizasyon yöntemi ile dönüştürülmesi sağlanmıştır. Ayrıca bu adımda çalışmada kullanılan KGI değerleri Cal/cm^2'den W/m^2 dönüştürülmüştür.

Verilerin normalizasyon süreci tamamlandıktan sonra her bir ilin veri setinin denetimli öğrenme algoritmalarının yapısına uygun olarak eğitim ve test şeklinde ayrılması gerekmektedir. Güneş tabanlı önceki yayınlar incelendiğinde farklı yöntem ve tekniklerin kullanılarak veri setinin ayarlandığı görülmüştür. Eğitim ve test verilerinin ayrımında üç farklı yöntem değerlendirilmiştir. İlk yöntemde ilgili yılın 12 ayının 8 ayı eğitime, geri kalan 4 ayı test verisi olarak ayarlanmıştır. İkinci yöntemde seçilen 4 yıllık

veri seti modelin eğitiminde kullanılıp, geri kalan yıl modeli test etmek amacıyla kullanılmıştır. Son yöntemde yazılan bir program kodu ile gün içinde ölçülen saatlik verilerden rastgele seçilen verilerin, üç saati eğitim bir saati test olacak şekilde ayarlanarak, il bazında verilerin yaklaşık %75'i eğitim, %25'i test olacak şekilde düzenlenmiştir. Her üç yöntemin karşılaştırmasında sonuçların çok fazla etkilememesi planlandığından veri setleri benzer şekilde modellenmiştir. Çizelge 4.2'de eğitim ve test verisinin seçimine göre geliştirilen modellerinin hesaplanan tahmin sonuçları gösterilmiştir. Model performansları istatistiki olarak değerlendirildiğinde üçüncü yöntemce geliştirilen model daha iyi tahmin doğruluğuna ulaştığından, çalışmada bu yöntemin kullanılması kararlaştırılmıştır. Ayrıca, eğitim ve test verilerinin yıl ve ay aralıkları değiştirildiğinde her defasında farklı tahmin sonuçlarına ulaşıldığı gözlemlenmiştir. Veri setlerinin rastgele saatlik olarak ayarlanması durumunda homojen bir dağılım sağlanarak tahmin modellerinin veri seçim aralıklarından etkilenmemesi sağlanmıştır.

Çizelge 4.2. Eğitim ve test verilerinin seçimine göre tahmin performansları

Yıllık				Aylık				Rastgele			
Eğitim	**Test**	**R^2**	**MSE**	**Eğitim (Ay No)**	**Test (Ay No)**	**R^2**	**MSE**	**Eğitim**	**Test**	**R^2**	**MSE**
2002-2005	2006	0.9206	0.0059	1-8	9-12	0.9102	0.0079				
2003-2006	2002	0.9214	0.0066	5-12	1-4	0.9302	0.0062				
2004, 2005, 2006, 2002	2003	0.9107	0.0075	1-4, 9-12	5-8	0.9165	0.0078	Gün Tabanlı Saatlik	Gün Tabanlı Saatlik	0.9534	0.0035
2002, 2003, 2005, 2006	2004	0.9165	0.0071	3-10	1-2, 11-12	0.9398	0.0053				
2002, 2003, 2004, 2005	2005	0,9369	0.0055	2-9	10-12, 1	0.9409	0.0043				

Çizelge 4.3'te Akdeniz bölgesinde çalışma alanı olarak seçilen sekiz ilin, veri ön işlem süreci sonunda ML tabanlı tahmin modellerinde kullanılacak şekilde düzenlenen eğitim ve test verilerinin nihai durumu ve temel özellikleri gösterilmektedir. Geliştirilen tahmin modellerinde 9 değişken girdi verisi, 1 değişken çıktı verisi olarak düzenlenerek toplam 10 farklı veri özelliğinden yararlanılmıştır. Her bir il için ML modellerin eğitiminde 17,584, testinde 5,858 adet rastgele seçilen saatlik veriden faydalanılmıştır.

Çizelge 4.3. Veri ön işleme süreci sonunda her bir il için düzenlenen veri setinin nihai özellikleri

	Meteorolojik ve Kategorik Tipte		Zaman Aralığı
	Eğitim Verileri	**Test Verileri**	
Toplam	17,584	5,858	2002-2006
Oran	±%75	±%25	
Seçim Tipi	Rastgele	Rastgele	
Girdi Verileri	Yıl, Ay, Gün, Saat, S, B, N, RH, GS		
Çıktı Verisi	KGI		

4.2. Girdi Verilerinden Optimal Özelliklerin Seçim Süreçleri

Günümüzde güneş tabanlı çalışmalarda olduğu gibi birçok alanda da veri miktarları ve türü göreceli olarak artmıştır. Eldeki tüm veri değişkenlerini kullanmaya çalışmak tahmin modellerinde her zaman üst düzey bir başarı getirmeyebilir. Bu nedenle ML'sine dayalı bir çalışma yapılmadan önce mutlaka kullanılacak veri özellikleri üzerinde birtakım işlemlerin yapılması gereklidir. Çünkü ML algoritmaları daha önce ifade edildiği üzere veri odaklı yapılar olarak bilinir ve bu temelde çalışırlar. Modelleme sırasında seçilen algoritmayı doğru, düzgün ve amaca uygun özellikler ile beslemek model başarısı için en önemli süreçlerdendir. ML tabanlı algoritmalarda sadece iyi düzenlenmiş verilere sahip olmak yetmemekte, aynı zamanda hedef çıktı ile uyumlu anlamlı özelliklere ve kullanışlı bir ölçeğe sahip olmak da gerekmektedir (Witten ve Frank, 2005). Bu nedenle, ML algoritmaları ile seçilecek en uygun veriler arasında çok kuvvetli bir ilişki bulunmaktadır. Geliştirilen modelin performansı üzerinde etkili olan bu ilişki genel olarak birçok çalışmada önemsenmemekte veya kısmen üzerinde durulmaktadır. Kullanılan algoritmada değişkenler, tahmin performansını pozitif veya negatif yönde etkileme potansiyeline sahip olan en önemli parametrelerden birisi olarak bilinmektedir. Bu nedenle veri seti içinden hedef çıktı üzerinde en etkili olan veri gruplarının belirlenmesi gereklidir. Optimal özelliklerin seçimi, modelleme adımını kolaylaştırmakta, aşırı uyumu engellemekte, daha hızlı ve performans gücü yüksek etkili modellerin geliştirilmesini sağlamaktadır. Hedeflenen başarı performansına çok daha çabuk ulaşılması sağlanmaktadır. Tersi durumda aynı performans seviyesine ulaşmak için çok daha karmaşık, uğraştırıcı, birden fazla modelin geliştirilmesi gerekecektir (Zheng ve Casari, 2018).

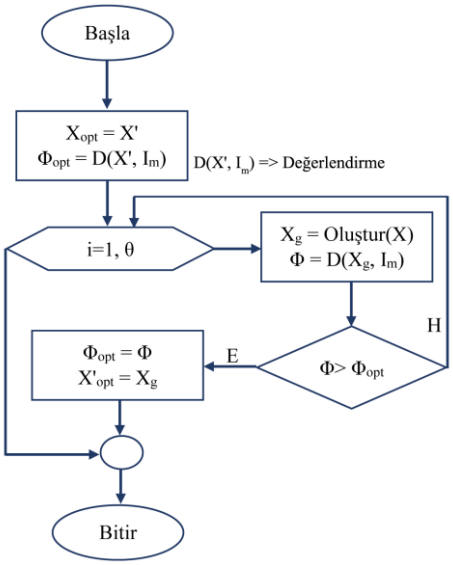
Özellik (değişken) denilen yapı, geliştirilen modelde veri görünümünün sayısal olarak temsil edilmesidir ve ML mimari yapısında veriler ile geliştirilen modeller

arasında bulunmaktadır. Özellik seçimi veya mühendisliği veri madenciliği içerisinde var olan özel bir alan olarak bilinmektedir. Veri Madenciliği temelde ML algoritmalarına dayandığından, tahmin modellerinin geliştirilmesinde uygulanan bu yöntemler, veri setinde var olan özellikleri ayıklama ve bunları ML modellerine uygun biçimlere uyarlama süreçleridir. Seçilen doğru özellikler modellemenin zorluğunu kolaylaştırmakta ve modellerin daha yüksek kalitede sonuçlar vermesini sağlamaktadır. Kapsamlı, detaylı ve zor bir süreç olmakla birlikte modelleme aşamasında birçok kez geri dönülmesi gerekebilmektedir (Zheng ve Casari, 2018).

KGI üzerinde etkili olan en uygun özellikleri belirlemek amacıyla açık kaynak kodlu WEKA 3.8.2 (Waikato Environment for Knowledge Analysis) programından faydalanılmıştır. Bu program gelişmiş veri madenciliği ve makine öğrenme tabanlı açık kaynak kodlu bir yazılımdır. Waikato üniversitesi tarafından JAVA programlama dili kullanılarak geliştirilmiştir. Programda hedef çıktı üzerinde en etkili verileri belirlemek amacıyla çok gelişmiş bir özellik seçim modülü bulunmaktadır. Modül, korelesyon temelli, sarmalayıcı ve filtre yaklaşımlarına dayanan üç özellik seçim yöntemi sunmaktadır (Witten ve ark., 1999). Bunlardan en sık bilinen ve kullanılanı filtre ve sarmalayıcı yaklaşımlardır. Filtre yaklaşımı ML algoritmasından bağımsız, hesaplama açısından basit, hızlı ve ölçeklenebilir bir yöntemdir. Şekil 4.2'de genel bir filtre yaklaşımının çalışma prensibi akış şeması olarak verilmiştir. Akış şemasında X, herhangi bir veri setindeki özellik grubunu, X' (ϕ, boş, herhangi bir özellik grubu) algoritmanın ilk aldığı başlangıç alt küme özelliklerini temsil etmektedir. I_m bağımsız bir ölçüt olarak oluşturulan her bir X_g alt kümesini değerlendirmekte ve daha önce oluşturulan en uygun alt kümeyle karşılaştırmaktadır. Durdurma kriteri olan θ değeri karşılanıncaya kadar optimal özelliklerin aranmasına devam edilmekte ve süreç sonunda X'_{opt} özellik alt kümesi seçilmektedir (Kumar ve Minz, 2014).

Sarmalayıcı yaklaşım ise veri setinde çıktıya etki eden en önemli değişkenleri bulmak için özellik alt küme alanında sınıflandırma tekniklerini kullanarak arama yapmaktadır. Filtre yaklaşımına benzer bir algoritmik yapısı olmakla birlikte en önemli fark, I_m bağımsız ölçütü yerine herhangi bir ML algoritmasının kullanılmasıdır. Diğer tüm yapısal özellikler Şekil 4.2'deki akış şemasına benzemektedir. Özellik alt kümesine eklenen ve çıkarılan her bir özellik için bir ML algoritmasının tahmin kesinliğinden faydalanılmaktadır. Bu nedenle seçilen özellik ile ML modeli arasında kuvvetli bir ilişki bulunmaktadır. Genel kanı odur ki sarmalayıcı yöntemler kullanılarak seçilen özellikler ile geliştirilen modellerin performansı, filtre yöntemleri ile seçilen özellikler

kullanılarak geliştirilen modellerden daha iyi tahmin sonucuna ulaşmaktadır (Kumar ve Minz, 2014).



Şekil 4.1. Filtre yaklaşımının algoritmik yapısının şematik gösterimi (Kumar ve Minz, 2014)

Özellik seçiminin en önemli etkisi girdi verilerini hedeflenen ML algoritması için optimize etmiş olmasıdır. Süreç sonunda bazı veriler, çıktı üzerinde pozitif etki yaparken, bazıları negatif etki yapabilmekte, bir kısmının ise hiçbir etkisi olmadığı belirlenebilmektedir. Bu sebeple modelleme sürecinden önce çıktı üzerinde en etkili veri özelliklerinin belirlenmesi, kullanılan verilerinin boyutsallığını azaltacak, yapılan tahminin yorumlanmasını kolaylaştıracak ve modelleme sürecini kısaltıp tahmin doğruluğunu arttıracaktır. WEKA programında, KGI'na en fazla etki eden değişkenleri seçmek için filtre ve sarmalayıcı yaklaşım temelli altı farklı özellik seçim fonksiyonu kullanılarak, her bir fonksiyonda birden fazla uygulama geliştirerek girdi değişkenlerinin güneş ışınımı üzerindeki etkisinin değerlendirilmesi amaçlanmıştır. Sarmalayıcı özellik seçim fonksiyonlarından Classifier Subset Evaluator (CSE), Wrapper Subset Evaluator (WSE) ve Classifier Attribute Evaluator (ClassAE) kullanılırken, filtre özellik seçim fonksiyonu olarak Corelation-based Future Selection Subset Evaluator (CfsSE), Corelation Attribute Evaluator (CorrAE) ve Relief Attribute Evaluator (RAE)'dan faydalanılmıştır. Birden fazla fonksiyonun seçilmesindeki amaç, güneş ışınım veri seti üzerinde farklı özellik seçim fonksiyonlarının etkinliğinin değerlendirilmesi için karşılaştırmalı bir çalışmanın yapılmasının sağlanmasıdır.

Özellik seçim sürecinde ayrıca birden fazla arama yönteminden faydalanılmıştır. Arama yöntemi, alt küme değerlendirmesine dayanarak olası özellik altkümelerinin arandığı yapısal bir yaklaşımdır. Rastgele ve kapsamlı arama gibi temel yöntemleri kullanılmakta ve özellik seçim fonksiyonunun türüne bağlı olarak seçilmektedir. Bazı seçim fonksiyonları birden fazla arama metodunu desteklerken, bazıları sadece birini desteklemektedir (Zheng ve Casari, 2018). En İyi Öncelikli Arama (BF), Evrimsel Arama (ES), Ateşböceği Arama (FS), Fil Arama (ELS), Karınca Arama (AE), Doğrusal İleri Seçim Arama (LFS), Açgözlü Adımsal Arama (GS) ve Ranker Arama algoritmaları kullanılmıştır. Bununla birlikte wrapper temelli özellik seçim fonksiyonlarında MLP, Doğrusal Regresyon (LR), Basit Doğrusal Regresyon (SLR), M5 Kuralları, M5P, Karar Tablosu (DT), RF, Toplamsal Regresyon (AR), Elastik-Net Regresyonu (ENR) ve K-NN Sınıflandırıcı (IBk) gibi on farklı ML algoritmasından sınıflayıcı olarak yararlanılmıştır.

Güneş ışınım verisine en fazla etki eden optimal girdi değişkenlerinin seçim sürecinde, toplamda altı farklı özellik seçim fonksiyonu kullanılarak 29 farklı uygulama geliştirilmiştir. Çizelge 4.4'te Mersin ili veri seti kullanılarak geliştirilen uygulamaların sonuçları verilmiştir. Çizelgede her bir uygulamanın sonucuna göre hesaplanan, KGI üzerinde en fazla etkisi olan özellikler (X) işareti ile gösterilirken, hiçbir etkisi olmayan veya çok az etkisi olan özellikler (-) işareti ile gösterilmiştir. Filtre yaklaşımı temelli 9 farklı uygulama geliştirilirken, sarmalayıcı yaklaşımı temelli 20 farklı uygulama geliştirilmiştir. Uygulamaların tamamında özellik seçim fonksiyonlarının yapısına göre farklı arama metotlarından faydalanılmış ve birden çok ML algoritmasından sınıflayıcı olarak yararlanılmıştır. Her bir uygulama için 10-Kat çapraz doğrulama yöntemi kullanılmıştır. Akdeniz bölgesinin diğer şehirleri için Mersin iline benzer çalışmalar yapılarak her bir uygulama için KGI en fazla etki eden özellikler belirlenmiştir.

WEKA programında, BF arama yöntemi kullanılarak CfsSE özellik seçim fonksiyonu temelinde geliştirilen bir özellik seçim uygulamasının ekran çıktısı Çizelge 4.5'te verilmiştir. Çizelgede, yıl, sıcaklık, rüzgâr hızı ve güneşlenme süresi verilerinin KGI üzerinde en etkili veriler olduğu, diğer verilerin hiçbir etkisinin olmadığı anlaşılmaktadır. Bazı özellik seçim fonksiyonlarında birden fazla uygulama geliştirilerek çıktı sonuçlarına göre KGI'na en fazla etki eden değişkenler belirlenerek gruplandırılmıştır. Birden fazla uygulama geliştirilen özellik seçim fonksiyonlarında nihai verilerin seçiminde toplam uygulamanın en az yarısında ilgili özelliğin seçilmesinin gerekliliği temel alınmıştır.

Çizelge 4.4. Mersin ili veri setine uygulanan 29 farklı uygulamanın seçim sonuçları

Uyg. Sayısı	Yaklaşım	Özellik Seçim Fonk.	Arama Metodu	ML Alg.	Seçilen Özellikler								
					Yıl	Ay	Gün	Saat	B	S	N	RH	GS
1	Filtre	CfsSE	BF	-	X	-	-	-	-	X	-	X	X
2	Filtre	CfsSE	ES	-	X	-	-	-	-	X	-	X	X
3	Filtre	CfsSE	GS	-	X	-	-	-	-	X	-	X	X
4	Filtre	CfsSE	FS	-	X	-	-	-	-	X	-	X	X
5	Filtre	CfsSE	ELS	-	X	-	-	-	-	X	-	X	X
6	Filtre	CfsSE	AS	-	X	-	-	-	-	X	-	X	X
7	Filtre	CfsSE	LFS	-	X	-	-	-	-	X	-	X	X
8	Sarmalayıcı	ClassAE	Ranker	DT	-	X	-	X	X	X	X	X	X
9	Sarmalayıcı	ClassAE	Ranker	ENR	-	-	-	-	X	X	-	X	X
10	Sarmalayıcı	ClassAE	Ranker	IBK	-	X	-	X	-	X	X	X	X
11	Sarmalayıcı	ClassAE	Ranker	M5	-	X	-	X	-	X	X	X	X
12	Sarmalayıcı	ClassAE	Ranker	MLP	-	-	-	X	-	X	-	-	X
13	Sarmalayıcı	ClassAE	Ranker	RF	-	X	-	X	-	X	X	X	X
14	Sarmalayıcı	ClassAE	Ranker	SLR	-	-	-	X	X	X	-	X	X
15	Sarmalayıcı	CSE	BF	DT	-	X	-	X	-	-	-	-	X
16	Sarmalayıcı	CSE	BF	ENR	X	X	-	X	-	X	X	X	X
17	Sarmalayıcı	CSE	BF	M5	X	X	X	X	X	X	-	-	X
18	Sarmalayıcı	CSE	BF	M5P	X	X	X	X	-	X	-	-	X
19	Sarmalayıcı	CSE	BF	MLP	X	X	X	X	X	X	X	X	X
20	Sarmalayıcı	CSE	BF	AR	-	X	-	X	-	-	-	-	X
21	Sarmalayıcı	CSE	BF	SLR	-	-	-	-	-	-	-	-	X
22	Filtre	CorrAE	Ranker	-	-	-	-	-	-	X	-	X	X
23	Filtre	RAE	Ranker	-	-	X	-	X	-	X	-	X	X
24	Sarmalayıcı	WSE	GS	AR	X	-	-	-	-	-	-	X	X
25	Sarmalayıcı	WSE	GS	DT	-	-	-	X	-	-	-	-	X
26	Sarmalayıcı	WSE	GS	ENR	X	X	-	X	-	X	X	X	X
27	Sarmalayıcı	WSE	GS	M5P	X	X	X	X	-	X	-	-	X
28	Sarmalayıcı	WSE	GS	MLP	X	X	X	X	X	-	X	X	X
29	Sarmalayıcı	WSE	GS	SLR	-	-	-	-	-	-	-	-	X

Çizelge 4.5. Filtre yaklaşımı temelli özellik seçim süreci sonunda geliştirilen uygulamanın ekran çıktısı

Evaluator:	weka.attributeSelection.CfsSubsetEval -P 1 -E 1
Search:	weka.attributeSelection.BestFirst -D 1 -N 5
Instances:	17,584
Attributes:	10 (Yıl, Ay, Gün, Saat, B, S, N, RH, GS, KGI)
Evaluation mode:	10-fold cross-validation
number of folds (%)	**attribute**
10(100 %)	1 Yıl
0(0 %)	2 Ay
0(0 %)	3 Gün
0(0 %)	4 Saat
0(0 %)	5 B
10(100 %)	6 S
0(0 %)	7 N
10(100 %)	8 RH
10(100 %)	9 GS

Akdeniz bölgesi illerinin her biri için seçilen özellik seçim fonksiyonları temelinde altı farklı veri grubu oluşturulmuştur. Model geliştirilen her bir ilin uygulanan seçim fonksiyonlarına göre KGI'na en fazla etki eden girdi değişkenleri Çizelge 4.6'da

görülmektedir. Birden fazla seçim uygulaması yapılan işlemlerde, en etkili özelliklerin seçimi, uygulama sayısı ile seçilen değişkenlerin etki toplamları değerlendirilerek belirlenmiştir. İşlemler sonucunda her bir özelliğin KGI üzerinde etki seviyesi negatif, nötr veya az ise ilgili özellik fonksiyonun seçim grubuna dahil edilmemiştir.

Çizelge 4.6. Akdeniz Bölgesi illerinin özellik seçim fonksiyonlarına göre seçilen veri özellikleri

		ADANA		**ANTALYA**		**BURDUR**	
No	**Özellik Seçim Fonksiyonu**	**Seçilen Özellikler**	**Toplam**	**Seçilen Özellikler**	**Toplam**	**Seçilen Özellikler**	**Toplam**
1	CfsSE	Yıl, S, N, RH, GS	5	S, RH, GS	3	S, RH, GS	3
2	ClassAE	Ay, Saat, S, N, GS	5	Yıl, Ay, Saat, B, S, N, RH, GS	8	Ay, Saat, B, S, N, RH, GS	7
3	CSE	**Ay, Saat, S, GS**	**4**	**Yıl, Ay, Gün, Saat, GS**	**5**	**Yıl, Ay, Gün, Saat, GS**	**5**
4	CorrAE	Gün, S, RH, GS	4	Yıl, Gün, Saat, S, RH, GS	6	Gün, S, RH, GS	4
5	RAE	Ay, Saat, S, N, RH, GS	6	Ay, Saat, S, RH, GS	5	Ay, Saat, S, GS	4
6	WSE	**Ay, Saat, S, GS**	**4**	**Yıl, Ay, Gün, Saat, GS**	**5**	**Yıl, Ay, Gün, Saat, GS**	**5**
		HATAY		**ISPARTA**		**KAHRAMANMARAŞ**	
No	**Özellik Seçim Fonksiyonu**	**Seçilen Özellikler**	**Toplam**	**Seçilen Özellikler**	**Toplam**	**Seçilen Özellikler**	**Toplam**
1	CfsSE	Ay, N, RH, GS	4	Yıl, N, GS	3	Yıl, S, N, RH, GS	5
2	ClassAE	Ay, Saat, B, S, N, RH, GS	7	Yıl, Ay, Saat, B, S, N, RH, GS	8	Ay, Saat, S, N, GS	5
3	CSE	Yıl, Ay, Gün, Saat, S, RH, GS	7	**Yıl, Ay, Gün, Saat, B, S, GS**	**7**	**Ay, Saat, S, GS**	**4**
4	CorrAE	Gün, Saat, S, RH, GS	5	Gün, S, RH, GS	4	Gün, S, RH, GS	4
5	RAE	Ay, Saat, S, GS	4	Ay, Saat, GS	3	**Ay, Saat, S, GS**	**4**
6	WSE	Ay, Saat, GS	3	**Yıl, Ay, Gün, Saat, B, S, GS**	**7**	Ay, Saat, S, N, RH, GS	6
		MERSİN		**OSMANİYE**			
No	**Özellik Seçim Fonksiyonu**	**Seçilen Özellikler**	**Toplam**	**Seçilen Özellikler**	**Toplam**		
1	CfsSE	Yıl, S, RH, GS	4	S, N, GS	3		
2	ClassAE	Ay, Saat, S, N, RH, GS	6	Ay, Saat, B, S, N, RH, GS	7		
3	CSE	**Yıl, Ay, Saat, GS**	**4**	**Yıl, Ay, Gün, Saat, S, GS**	**6**		
4	CorrAE	S, RH, GS	3	Gün, S, RH, GS	4		
5	RAE	Ay, Saat, S, RH, GS	5	Ay, Saat, S, GS	4		
6	WSE	**Yıl, Ay, Saat, GS**	**4**	**Yıl, Ay, Gün, Saat, S, GS**	**6**		

Çizelge 4.6'da bazı özellik seçim fonksiyonları için hesaplanan girdi değişkenleri ve sayısı birbirlerine benzer çıkmıştır. Adana, Antalya, Burdur, Isparta

Mersin ve Osmaniye illerinin, CSE ve WSE özellik seçim fonksiyonları için hesaplanan özellikler birbirlerine benzer çıkarken, Kahramanmaraş ilinde CSE ve RAE özellik seçim fonksiyonları sonuçlarının aynı bulunduğu görülmektedir. Yalnız Hatay ilinin veri setinde yapılan özellik seçim süreçlerinde, hiçbir özellik seçim fonksiyonun sonuçları birbirlerine benzememektedir. Akdeniz bölgesinde çalışma alanı olarak seçilen her bir il için, özellik seçim süreçleri sonucunda ML algoritmaları ile geliştirilecek tahmin modellerinde kullanılmak üzere oluşturulan nihai veri grupları ve özellik sayıları Çizelge 4.7'de verilmiştir.

Çizelge 4.7. Akdeniz Bölgesi illerinin tahmin modellerinde kullanılmak üzere oluşturulan nihai girdi veri grupları ve özellik sayıları

Oluşturulan Gruplar	ADANA		ANTALYA		BURDUR		ISPARTA	
	Özellik Seçim Fonk.	Özellik Sayısı	Özellik Seçim Fonk.	Özellik Sayısı	Özellik Seçim Fonk.	Özellik Sayısı	Özellik Seçim Fonk.	Özellik Sayısı
GR1	CfsSE	5	CfsSE	3	CfsSE	3	CfsSE	3
GR2	ClassAE	5	ClassAE	8	ClassAE	7	ClassAE	8
GR3	CSE-WSE	4	CSE-WSE	5	CSE-WSE	5	CSE-WSE	7
GR4	CorrAE	4	CorrAE	6	CorrAE	4	CorrAE	4
GR5	RAE	6	RAE	5	RAE	4	RAE	3
Oluşturulan Gruplar	**K. MARAŞ**		**MERSİN**		**OSMANİYE**		**HATAY**	
	Özellik Seçim Fonk.	**Özellik Sayısı**	**Özellik Seçim Fonk.**	**Özellik Sayısı**	**Özellik Seçim Fonk.**	**Özellik Sayısı**	**Özellik Seçim Fonk.**	**Özellik Sayısı**
GR1	CfsSE	5	CfsSE	4	CfsSE	3	CfsSE	4
GR2	ClassAE	5	ClassAE	6	ClassAE	7	ClassAE	7
GR3	CSE-RAE	4	CSE-WSE	4	CSE-WSE	6	CSE	7
GR4	CorrAE	4	CorrAE	3	CorrAE	4	CorrAE	5
GR5	WSE	6	RAE	5	RAE	4	RAE	4
GR6	-	-	-	-	-	-	WSE	3

Girdi veri grupları kullanılarak seçilen ML algoritmalarına göre KGI tahmin modellerinin geliştirilmesi planlanmıştır. Normal şartlarda belirlenen 9 girdi değişkeni için $2^9 - 1 = 511$ farklı özellik kombinasyonuna dayanan ML tabanlı tahmin modelleri geliştirmek, analiz etmek ve performanslarını karşılaştırmak gereklidir. Bu neredeyse imkânsız zor bir süreçtir. Bu nedenle güneş ışınım tahmin çalışmalarında tercih edilmeyen özellik seçim süreçleri sayesinde verilerin rastgele bir şekilde seçilmesinin önüne geçilerek, verilerin birbirleri arasında ilişki gücünün tahmin performanslarına doğru bir şekilde yansıması sağlanmıştır. Ayrıca geliştirilen modellerin yorumlanmasına, hızına ve performansına önemli katkılar sağladığı düşünüldüğünde özellik seçim süreçlerinin karşılaştırmalı bir değerlendirmesi de sunulmuştur.

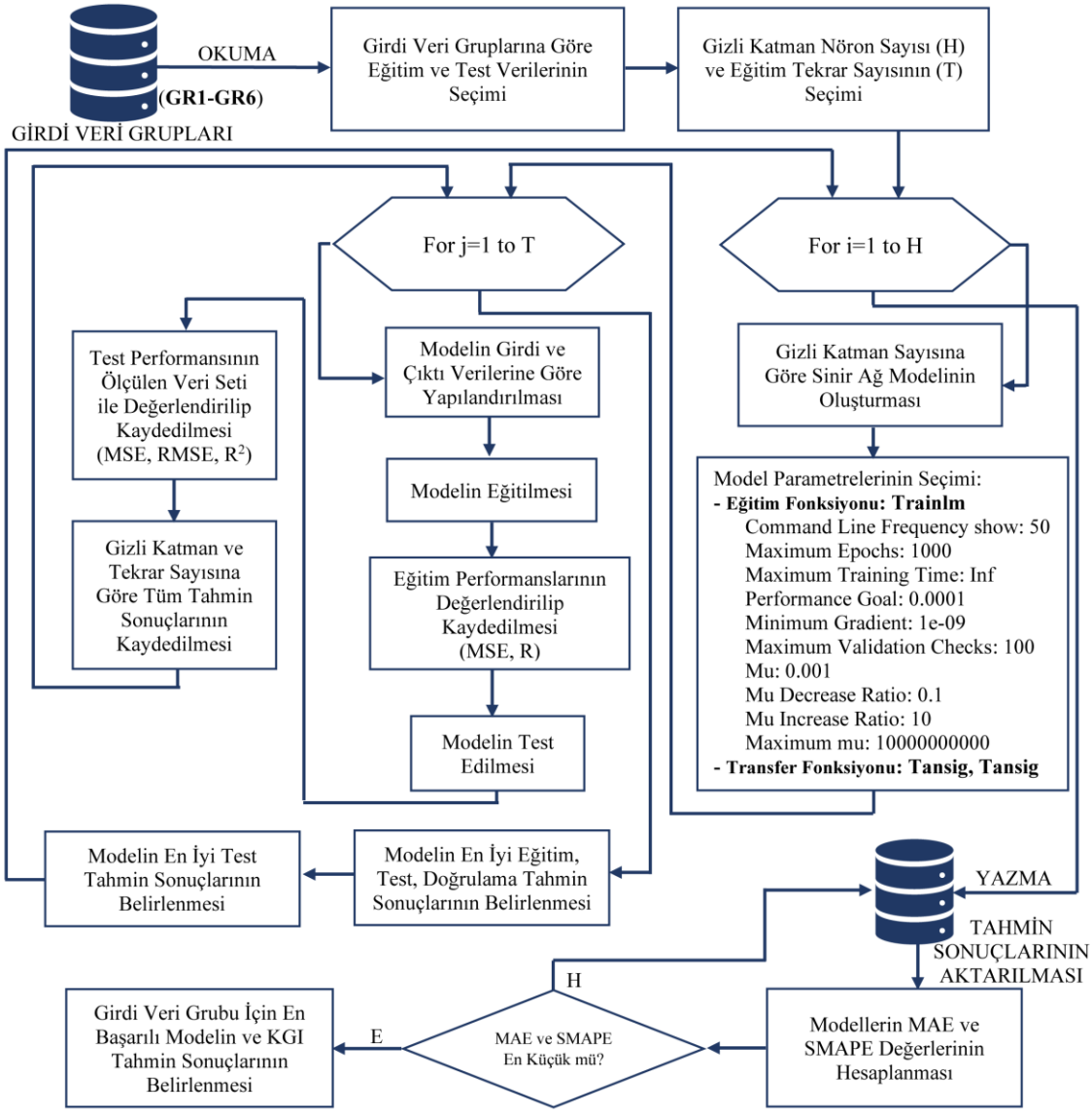
5. GELİŞTİRİLEN MODELLERİN BULGULARI VE TARTIŞMA

Bu bölümde veri madenciliği süreçleri kullanılarak düzenlenen ve seçilen girdi veri grupları ile Akdeniz bölgesinde herhangi bir yerde kurulacak güneş enerjisi santrali için KGI'nın tahmin edilmesi amacıyla geliştirilen ML tabanlı modellerin analizleri yapılmıştır. Bölgede seçilen her bir ilin özellik seçim gruplarına göre tahmin modelleri geliştirilmiştir. İlk olarak ÇKİBSA ML algoritması kullanılarak modeller geliştirilmiş ve her bir ilin performans sonuçları istatistiki olarak analiz edilerek geliştirilen modellerin veri tahmin ilişkisi değerlendirilmiştir. Ardından klasik DVR, LibSVM, M5 Kuralları, K-NN ML algoritmaları kullanılarak seçilen belirli iller için geliştirilen modeller karşılaştırılmıştır. Bu aşamada veriden ziyade algoritmaların performansı hedef alınmıştır. Güneş ışınım çalışmalarında kullanılan tüm tahmin yaklaşımları bu bölümde değerlendirilmiştir. VM süreçleri ile ML algoritmalarının hibrit bir şekilde kullanılmasının tahmin sonuçlarına etkisi gösterilmiştir. Tartışma bölümünde ise ML algoritmaların modelleme sonuçları genel olarak analiz edilmiştir.

5.1. ÇKİBSA Algoritması ile Tahmin Modellerinin Geliştirilmesi

Güneş tabanlı veri-tahmin ilişkisinin modellenmesinde en fazla yararlanılan ML algoritmalarının YSA mimarilerine dayalı algoritmalar olduğu görülmüştür. Yaygın kullanımın en önemli sebepleri arasında veri-tabanlı hassas algoritmalar olması yatmaktadır. Ayrıca seçilen veriye uygun doğru modelleme yapıldığında oldukça yüksek performans seviyelerine ulaşıldığı bilinmektedir (Long ve ark., 2014). Bu nedenle hem özellik seçim süreci sonunda belirlenen veri gruplarının etki seviyelerini değerlendirmek hem de KGI'nı yüksek doğrulukta tahmin edebilmek amacıyla ÇKİBSA sinir ağı mimarisine dayalı algoritmadan faydalanılmıştır. Şekil 5.1'de Matlab yazılımında ilgili algoritmanın mimari yapısı temel alınarak tasarlanan program kodunun işlem basamakları gösterilmiştir. Yazılan kod çalıştığında Akdeniz bölgesi illerinin VM süreçleri sonunda seçilen veri gruplarından yararlanılarak tahmin modelleri geliştirilmiştir. Adana, Antalya, Burdur, Isparta, Kahramanmaraş, Mersin ve Osmaniye illeri için özellik seçimine dayalı 5 farklı gruptan (GR1-GR5) faydalanılırken, sadece Hatay ili için 6 farklı gruptan (GR1-GR6) yararlanılmıştır. Ayrıca her bir il için özellik seçimi sonunda belirlenen veri gruplarının tahmin doğruluklarını değerlendirmek amacıyla tüm girdi değişkelerinin dahil olduğu bir başka grup veri havuzuna dahil

edilmiştir. Böylece geliştirilen modellerde GR6, Adana, Antalya, Burdur, Isparta, Kahramanmaraş, Mersin ve Osmaniye illeri için oluşturulan tüm girdi değişkenlerinin dahil olduğu veri grubunu temsil ederken, sadece Hatay ili için GR7 tüm girdi özellikleri yerine kullanılmıştır.



Şekil 5.1. ÇKİBSA algoritmasına göre geliştirilen modellerin akış şeması

Tüm YSA mimarilerinde geliştirilen modellerin performansında en belirleyici olan unsurların gizli katman, gizli katmandaki nöron ve girdi sayılarının belirlenmesi olduğu bilinmektedir. Ancak yapılan çalışmalarda bu sayıları optimum düzeyde seçmenin etkili bir yöntemi bulunmamaktadır. Gizli katmanlar, ağa genelleme yeteneği sağladığından bir veya bazen iki gizli katmana sahip sinir ağları yaygın olarak

kullanılmaktadır. Gizli katman sayısını arttırmak, hesaplama süresini uzatıp aşırı uyuma yol açtığından dikkatli olunması gerekmektedir. Benzer şekilde gizli katmandaki nöron sayılarını maksimum ağ performansına göre seçmenin de kesinleşmiş bir kuralı bulunmamaktadır. Nöronları seçmede sabit, yapıcı ve yıkıcı olmak üzere üç farklı yöntem kullanılmaktadır. Sabit yöntemde, her biri farklı sayıda gizli nörona sahip sinir ağları oluşturularak eğitilmekte ve test edilmektedir. Gizli nöronların artışı hesaplama süresi ve kaynağına bağlı olarak birer, ikişer veya daha fazla sayıda arttırılabilmektedir. Her bir nöronun için geliştirilen modelin ağ performansı değerlendirilerek en başarılı model bulunmaktadır. Kullanılan bu yöntem zaman alıcı olmakla birlikte en iyi performans seviyesine ulaşmada kesin sonucu verebilmektedir. Yapıcı yöntem, ağ performansı bozulmaya başlayıncaya kadar, eğitim sırasında gizli nöronların eklenmesi, yıkıcı yöntem ise gizli nöronların çıkarılması mantığına dayanmaktadır. Her iki yöntemde de sabit yöntemde olduğu gibi her bir gizli nöron için ayrı ayrı sinir ağı modelleri oluşturmak gerekmez. Fakat birçok YSA tabanlı yazılım programları eğitim sırasında gizli nöronların değiştirilmesini desteklemediğinden yapıcı ve yıkıcı yöntemler, çok bilinen ve kullanılan yöntemler olarak bilinmemektedirler (Reed ve Marks, 1999).

Modelleme çalışmalarında hesaplama süresini kısaltmak ve hesaplama kaynağını verimli kullanmak adına Intel Xeon 2.40 GHz (16 çekirdekli) çift işlemcili, 16 GB ram ve 640 CUDA 4 GB 128 bit ekran kartı olan bir iş istasyonundan faydalanılmıştır. Geliştirilen sinir ağlarında, gizli katman sayısı arttırıldığında hesaplama süresi oldukça uzadığından ve belirli bir performans artışı sağlanamadığı görüldüğünden bir gizli katmanın kullanılması kararlaştırılmıştır. Her bir ilin en başarılı modelinin bulunması için gizli katmanlardaki optimal nöron sayıların belirlenmesinde sabit yöntemin kullanılması kararlaştırılmıştır. ÇKİBSA mimarisine göre geliştirilen program kodu maksimum gizli katman nöron sayısı ve eğitim tekrar sayısının programın başında girilmesini isteyen esnek bir yapıda tasarlanmıştır. Ancak yapılan deneme çalışmalarında, gizli katman sayısı 50'nin üzerine çıkarıldığında modelleme süresi oldukça uzamakta ve belirli bir performans artışı sağlanmayıp, geliştirilen modellerde aşırı uyum ile performans kayıplarının yaşandığı gözlemlenmiştir. Bu sebepten tüm illerde bir standardın sağlanması adına maksimum gizli katman sayısı 50 ile eğitim tekrarı 5 ile sınırlı tutulmuştur. Şekil 5.1'de görüleceği üzere, program çalıştığında girilen maksimum gizli katman nöron sayısına kadar, nöron sayıları birer birer arttırılarak, her bir nöron için modeller geliştirilmektedir. Tekrar sayısı, her bir model

için birden fazla eğitim prosesi görevini yerine getirmekte ve eğitim sonunda ilgili gizli nöron sayısı için en başarılı olan model hesaplanmaktadır.

Model performansları değerlendirilirken ağın tüm parametrelerinin sabit tutulması gerekmektedir. Geliştirilen sinir ağı modelinde herhangi bir parametrenin değiştirilmesi potansiyel olarak farklı bir hata göstergesine sahip yeni bir sinir ağının oluşturulması demek olduğundan, optimum sayıda gizli nöron seçimini gereksiz yere karmaşıklaşmaktadır (Reed ve Marks, 1999). Bu sebeple, modellerde optimum ağ parametrelerin belirlenmesinde gizli katman ve gizli nöron sayıları sabit tutularak farklı denemeler gerçekleştirilerek ağ performansları karşılaştırılmış, Şekil 5.1'deki model parametrelerinin geliştirilen tüm modellerde kullanılması kararlaştırılmıştır. Tasarlanan yazılım çalıştığında her bir ilin girdi veri grupları kullanılarak 1-50 nöron aralığında tahmin modelleri oluşturulmaktadır. Ardından her bir nöron için belirlenen ağ parametrelerine göre geliştirilen modeller yapılandırılmakta ve tekrar sayısı kadar eğitilmektedir. Test veri seti kullanılarak gerçek ölçülen veriler yardımıyla geliştirilen her bir modelin performansları istatistiki olarak hesaplanmaktadır. Bölge illerinin her bir girdi veri grubu için 250 farklı tahmin modeli geliştirilmekte ve performansları analiz edilmektedir. Toplamda her bir il için ortalama 1500 farklı ÇKİBSA modeli geliştirilmiştir. Süreç sonunda her girdi grubu için en başarılı eğitim ve test tahmin performansları istatistiki olarak belirlenmiş ve kaydedilmiştir. Akdeniz bölgesinde çalışma alanı olarak seçilen her bir ilin tasarlanan program kodu ile ÇKİBSA ML algoritması kullanılarak geliştirilen tahmin modelleri ayrı ayrı değerlendirilip karşılaştırılmıştır.

5.1.1. Adana ilinin tahmin bulguları

Adana, 35°-38° kuzey enlemleri ile 34°-46° doğu boylamları arasında yer almaktadır. Doğu Akdeniz'de bulunan ilin sahil uzunluğu 160 Km, yüz ölçümü 17,253 km^2 ve şehir merkezinin deniz seviyesinden yüksekliği ortalama 23 m'leri bulmaktadır. Yıllık sıcaklık ortalaması 19.2 °C, ortalama güneşlenme süresi 7.5 saat seviyelerinde hesaplanmıştır (MGM, 2020). Bölge yıl boyu güneş enerji potansiyeli bakımından yüksek bir potansiyele sahiptir. Adana ili için MGM'den talep edilen veriler 17,351 numaralı istasyondan toplanmıştır. Hedef ile konumlandırılan istasyon, 37.041 enlem, 35.344 boylam ve deniz seviyesinden 23 m yükseklikte faaliyet göstermektedir.

2002-2006 arası saatlik olarak ölçülen veriler VM süreçlerine sokularak düzenlenmiş ve her bir özellik için 23,442 adet veri il için kullanılabilir hale getirilerek özellik seçimine dayalı veri grupları oluşturulmuştur. ÇKİBSA algoritması ile KGI tahmin etmek için GR1-GR6 aralığındaki veri grupları kullanılmış ve her bir veri grubu için birden fazla farklı özelliklerden yararlanılarak tahmin modelleri geliştirilmiştir. Geliştirilen her bir sinir ağı modelinin eğitim, doğrulama ve test KGI tahminleri istatistiksel olarak değerlendirildiğinde, en başarılı ÇKİBSA tahmin modellerinin sonuçları Çizelge 5.1'de verilmiştir.

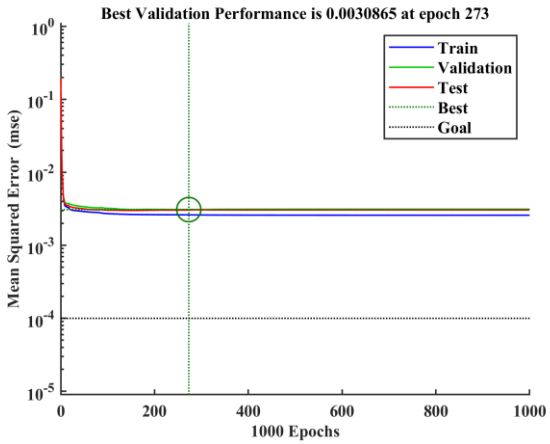
Çizelge 5.1. Adana ili veri gruplarına göre geliştirilen en başarılı modellerin eğitim ve test tahmin sonuçları

		Modellerin Eğitim Performansı						**Modellerin Test Performansı**			
Veri Grubu	**Gizli Katman Nöron Sayısı**	**Eğitim Tekrar Sayısı**	**Eğitim Verisi MSE**	**Test Verisi MSE**	**Doğrulama Verisi MSE**	**Model MSE**	**Model R**	**RMSE**	**MAE**	**SMAPE (%)**	**R^2**
GR1	40	2	0.0154	0.0158	0.0158	0.0165	0.8689	0.1300	0.1003	12.75	0.7385
GR2	33	5	0.0031	0.0034	0.0033	0.0035	0.9745	0.0591	0.0426	6.73	0.9460
GR3	***48***	***1***	***0.0026***	***0.0029***	***0.0027***	***0.0030***	***0.9790***	***0.0544***	***0.0386***	***6.15***	***0.9542***
GR4	28	3	0.0185	0.0191	0.0187	0.0188	0.8424	0.1407	0.1107	13.60	0.6937
GR5	43	1	0.0033	0.0034	0.0034	0.0034	0.9736	0.0597	0.0425	6.47	0.9448
GR6	45	2	0.0025	0.0029	0.0027	0.0029	0.9785	0.0553	0.0402	8.24	0.9527

Çizelgeden görüleceği üzere GR2, GR3 ve GR5 özellik seçim veri grupları ve GR6 tüm girdi değişkenleri kullanılarak geliştirilen modellerin eğitim performansları birbirlerine oldukça yakın hesaplanmıştır. GR2 veri grubunda ClassAE özellik seçim fonksiyonuna dayalı 5 özellik, GR3 veri grubunda CSE-WSE özellik seçim fonksiyonuna dayalı 4 özellik ve GR5 veri grubunda RAE özellik seçim fonksiyonuna dayalı 6 özellik kullanılarak tahmin modelleri geliştirilmiştir. Bu dört veri grubundan, GR3 veri grubuyla geliştirilen en başarılı modelin eğitim tahmin sonucu 48 gizli katman nöronu ile birinci tekrarda **MSE = 0.0030** ve **R = 0.9790** olarak hesaplanarak diğer grupların üzerinde bir performans seviyesine ulaşmıştır. GR6 veri grubuyla geliştirilen en başarılı modelin eğitim performansı ise 45 gizli katman nöronu ile ikinci tekrarda **MSE = 0.0029** ve **R = 0.9785** olarak bulunmuştur. Geliştirilen modellerin test tahmin performanslarının değerlendirilmesinde SMAPE, MAE, RMSE ve R^2 istatistiksel hata ölçüm ve analiz yöntemleri kullanılarak, oluşturulan veri gruplarının tahmin sonuçları değerlendirilmiştir. İstatistiksel hata ölçüm yöntemlerinde ölçüm ile tahmin arasındaki değer 0'a ne kadar yakın olursa, analiz yöntemlerinde ise 1'e ne kadar yakın olursa geliştirilen modellerin tahmin doğrulukları o oranda yüksek olmaktadır (Tasdemir ve

ark., 2018). GR3 ve GR6 modellerinin test verileri kullanarak en iyi tahmin sonuçları ölçülen gerçek değerlerle karşılaştırıldığında, GR3 için R^2, RMSE, MAE ve SMAPE değerleri sırasıyla **0.9542, 0.0544, 0.0386** ve **%6.15** sonuçlarına ulaşılırken, GR6 için sırasıyla **0.9527, 0.0553, 0.0402** ve **%8.24** olarak hesaplanmıştır. GR3 veri grubuyla geliştirilen en başarılı modelin eğitim ve test tahmin performansları diğer veri gruplarına nazaran daha doğru ve kesin bulunmuştur.

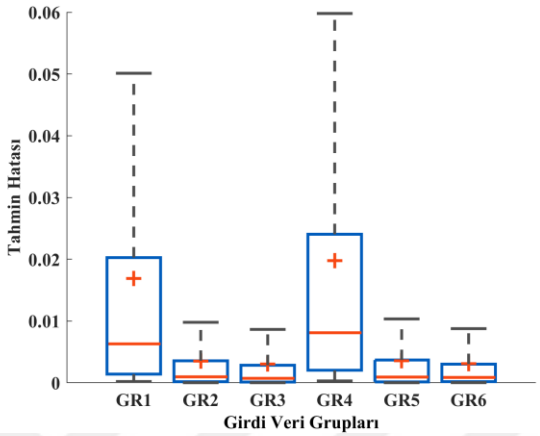
Eğitim sürecinde en iyi tahmin performansına ulaşan GR3 modelinin ağ performans grafiği Şekil 5.2'de gösterilmiştir. Bu model, 0 ile 1000 epok değerleri arasında eğitilmiştir. En başarılı performansa 273 epok sonucunda ulaşılmıştır. Geliştirilen modelin MSE tahmin sonucunun **0.0031** düzeylerinde olduğu görülmektedir.



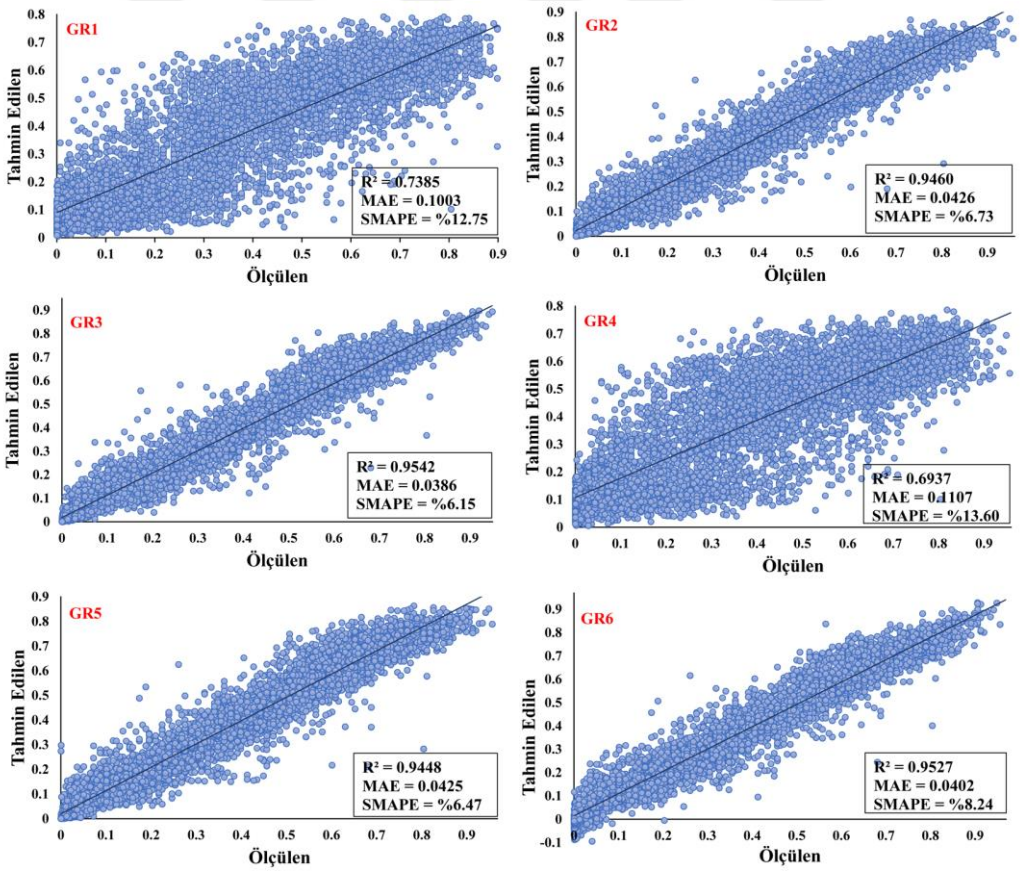
Şekil 5.2. Adana ili için geliştirilen en iyi modelin ÇKİBSA performans grafiği

Adana ilinin girdi veri grupları üzerinde yapılan tahmin çalışmalarında ölçülen ve tahmin edilen değerleri arasındaki hata kutu grafikleri Şekil 5.3'te verilmiştir. Bu grafiklerde her bir veri grubu için sistemin daha önce görmediği 5,858 adet ölçülen saatlik girdi verisine göre tahmin edilen değerler arasındaki istatistiksel hata ortalama ölçüm sonuçları görülmektedir. Hata oranları azaldıkça kutu grafiğin boyutu küçülmekte ve sıfıra yaklaşarak tahmin performansı en iyi olan modeli işaret etmektedir. Kutu grafiklerinde kırmızı çizgi ortancayı, '+' simgesi ortalamayı temsil etmektedir. Her bir veri grubu için geliştirilen en başarılı modellerin ölçülen ve tahmin edilen değerleri arasındaki dağılım grafikleri Şekil 5.4'te verilmiştir. Grafiklerden GR2, GR3, GR5 ve GR6 veri grupları için yüksek doğrulukta bir korelasyon seviyesine ulaşıldığı anlaşılmaktadır. Korelasyon katsayısı iki farklı veri seti arasındaki ilişkinin

yönü ve şiddeti hakkında bilgi veren güvenilir bir istatistiki analiz yöntemidir. Korelasyon katsayısının, 0.90-1.00 arasında olması tahmin ile gözlemlenen değerler arasında çok yüksek düzeyli bir ilişkinin varlığını göstermektedir (Asuero ve ark., 2006).

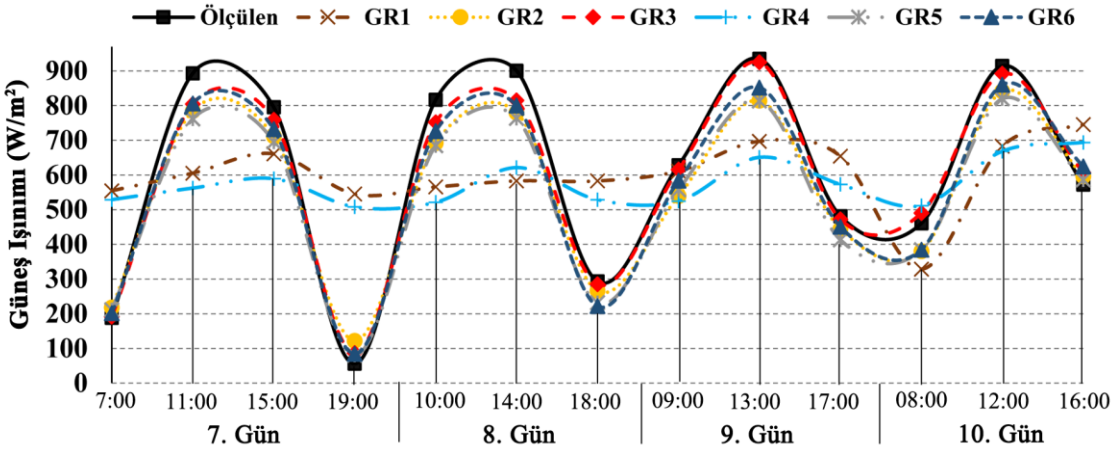


Şekil 5.3. Adana ili veri gruplarına göre ölçülen ve tahmin edilen değerler arasındaki hata kutu grafiği



Şekil 5.4. Adana ili test veri gruplarına göre geliştirilen ÇKİBSA modellerinin KGI tahmin dağılım grafikleri

Adana ili için özellik seçim fonksiyonlarına dayalı GR1-GR5 arası veri grupları ile geliştirilen ÇKİBSA tahmin modellerinden GR2, GR3 ve GR5 veri gruplarıyla yapılan tahminlerin hedef yerin güneş enerjisi potansiyelini iyi bir şekilde tahmin ettiği sonucuna varılabilir. Ancak, GR1 ve GR4 veri grubuyla geliştirilen tahmin modellerinin performansı oldukça düşük çıkmıştır. Meteorolojik ve kategorik girdi veri setine uygulanan özellik seçim fonksiyonlarından CfsSE ve CorrAE'nın en iyi girdi özelliklerini belirlemede yetersiz kaldığı sonucuna varılmıştır. En başarılı özellik seçim fonksiyonları ClassAE, CSE, WSE ve RAE'dir. GR6 veri grubu ile karşılaştırıldığında CSE-WSE özellik seçim fonksiyonlarına dayanan GR3 girdi veri grubu ile modellenen sinir ağının daha kesin tahmin doğruluğuna ulaştığı hesaplanmıştır. Hedef ilin girdi veri gruplarına göre geliştirilen en iyi modellerinin güneş ışınım tahminleri ile gerçek ölçüm değerlerinin karşılaştırmalı performans grafiği Şekil 5.5'te verilmiştir. Test girdi verilerinden rastgele seçilen 2003 yılının Mayıs ayının 7-10 günleri arası dört günlük saatlik girdi verileri kullanılmıştır. Her gün için rastgele seçilerek oluşturulan test verilerinin ilgili saat dilimlerinde ölçüm ve tahmin sonuçları karşılaştırılmış ve GR3 veri grubuyla geliştirilen tahmin modelinin diğer modellere kıyasla daha başarılı bir şekilde güneş ışınımını tahmin ettiği anlaşılmıştır. Modelleme çalışmaları sonunda, Adana ili için VM sürecine dayalı geliştirilen tahmin modellerinin KGI'nı başarılı bir şekilde tahmin ettiği sonucuna varılabilmektedir.



Şekil 5.5. Adana ili 2003 yılı Mayıs ayı dört günlük ölçülen ve tahmin edilen saatlik verilerin karşılaştırmalı performans grafiği

5.1.2. Antalya ilinin tahmin bulguları

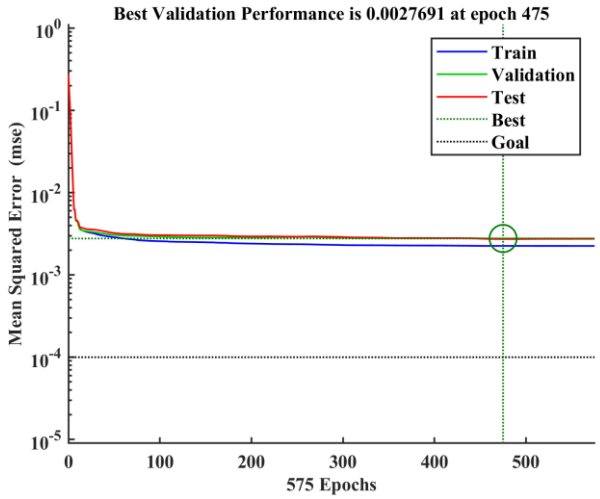
Akdeniz Bölgesi'nin batısında bulunan Antalya ilinin yüzölçümü 20,815 km^2 ile bölgenin %17.6'sını kapsamaktadır. Hedef il, 29°-32° doğu boylamları ile 36°-37° kuzey enlemleri arasında yer almaktadır. Yazın sıcaklık ortalaması 28-36 °C arasında, kışın ise 10-20 °C arasında değişmektedir. Yıllık sıcaklık ortalaması 18.8 °C, ortalama güneşlenme süresi 8.2 saat seviyelerinde hesaplanmıştır (MGM, 2020). Antalya ilinin meteorolojik verileri 17,302 numaralı istasyondan alınmıştır. Seçilen istasyonun enlemi 36.885, boylamı 30.683 ve deniz seviyesinden yüksekliği 47 m'dir.

İlin VM süreçleri sonunda düzenlenen GR1-GR6 arası girdi verileri kullanılarak 3 katmanlı mimari yapıya dayanan ÇKİBSA algoritması ile KGI tahmin etmek için 1500 farklı tahmin modeli geliştirilmiştir. Geliştirilen her bir sinir ağı modelinin tahmin performansları istatistiksel olarak değerlendirildiğinde veri grupları temelinde en başarılı modellerin tahmin sonuçları Çizelge 5.2'de gösterilmiştir. GR2 veri grubunda ClassAE özellik seçim fonksiyonuna dayalı 8 girdi değişkeni, GR3 veri grubunda CSE-WSE özellik seçim fonksiyonuna dayalı 5 girdi değişkeni ve GR6 veri grubunda tüm değişkenlerin kullanıldığı modellerin tahmin performansı daha üst düzeyde hesaplanmıştır. **GR2, GR3** ve **GR6** veri grupları ile geliştirilen en iyi modellerin R istatistik ölçeğine göre eğitim sonuçları sırasıyla **0.9775, 0.9793** ve **0.9769** değerleri ile birbirlerine oldukça yakın hesaplanmıştır. Bu veri gruplarının test tahmin performansları SMAPE, MAE, RMSE ve R^2 istatistiksel ölçeklerine göre **GR2** için sırasıyla **%7.14, 0.0367, 0.0533** ve **0.9515**, GR3 için sırasıyla **%7.13, 0.0353, 0.0516** ve **0.9546**, GR6 için sırasıyla **%7.20, 0.359, 0.531** ve **0.9519** değerleri bulunmuştur. Grupların test performansları birbirlerine oldukça yakın çıksa da GR3 veri grubuyla geliştirilen modelin tahmin performansı Antalya ili için SMAPE, MAE ve R^2 ölçeklerine göre güneş ışınımını başarılı bir şekilde tahmin ettiği sonucuna varılmıştır.

Çizelge 5.2. Antalya ili veri gruplarına göre geliştirilen en başarılı modellerin eğitim ve test tahmin sonuçları

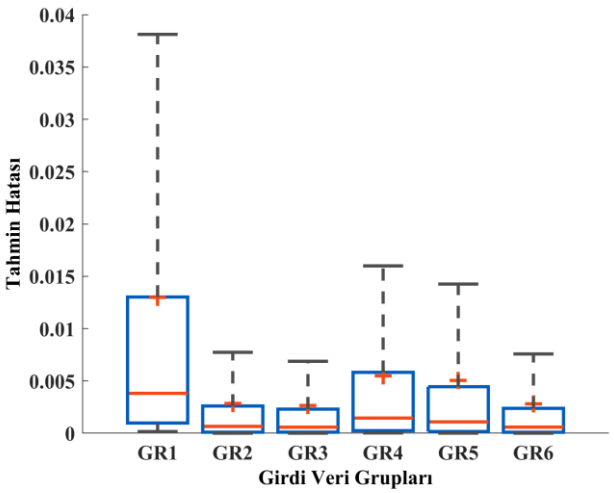
		Modellerin Eğitim Performansı						**Modellerin Test Performansı**			
Veri Grubu	**Gizli Katman Nöron Sayısı**	**Eğitim Tekrar Sayısı**	**Eğitim Verisi MSE**	**Test Verisi MSE**	**Doğrulama Verisi MSE**	**Model MSE**	**Model R**	**RMSE**	**MAE**	**SMAPE (%)**	**R^2**
GR1	27	4	0.0128	0.0124	0.0130	0.0131	0.8816	0.1140	0.0847	12.59	0.7780
GR2	42	3	0.0025	0.0027	0.0026	0.0026	0.9775	0.0533	0.0367	7.14	0.9515
GR3	***44***	***3***	***0.0023***	***0.0026***	***0.0024***	***0.0028***	***0.9793***	***0.0516***	***0.0353***	***7.13***	***0.9546***
GR4	50	2	0.0048	0.0057	0.0052	0.0055	0.9548	0.0739	0.0535	9.41	0.9068
GR5	30	4	0.0047	0.0048	0.0049	0.0049	0.9573	0.0710	0.0487	7.98	0.9140
GR6	48	4	0.0025	0.0029	0.0027	0.0030	0.9769	0.0531	0.0359	7.20	0.9519

VM süreci sonunda belirlenen GR3 veri grubu kullanılarak 44 gizli katman nöronlu eğitim modelinin 3. tekrar sonucundaki ağ performans grafiği Şekil 5.6'da gösterilmiştir. En başarılı performansa 475 epok sonucunda ulaşılmıştır. MSE tahmin performansının yaklaşık 0.0028 seviyelerinde olduğu görülmektedir.

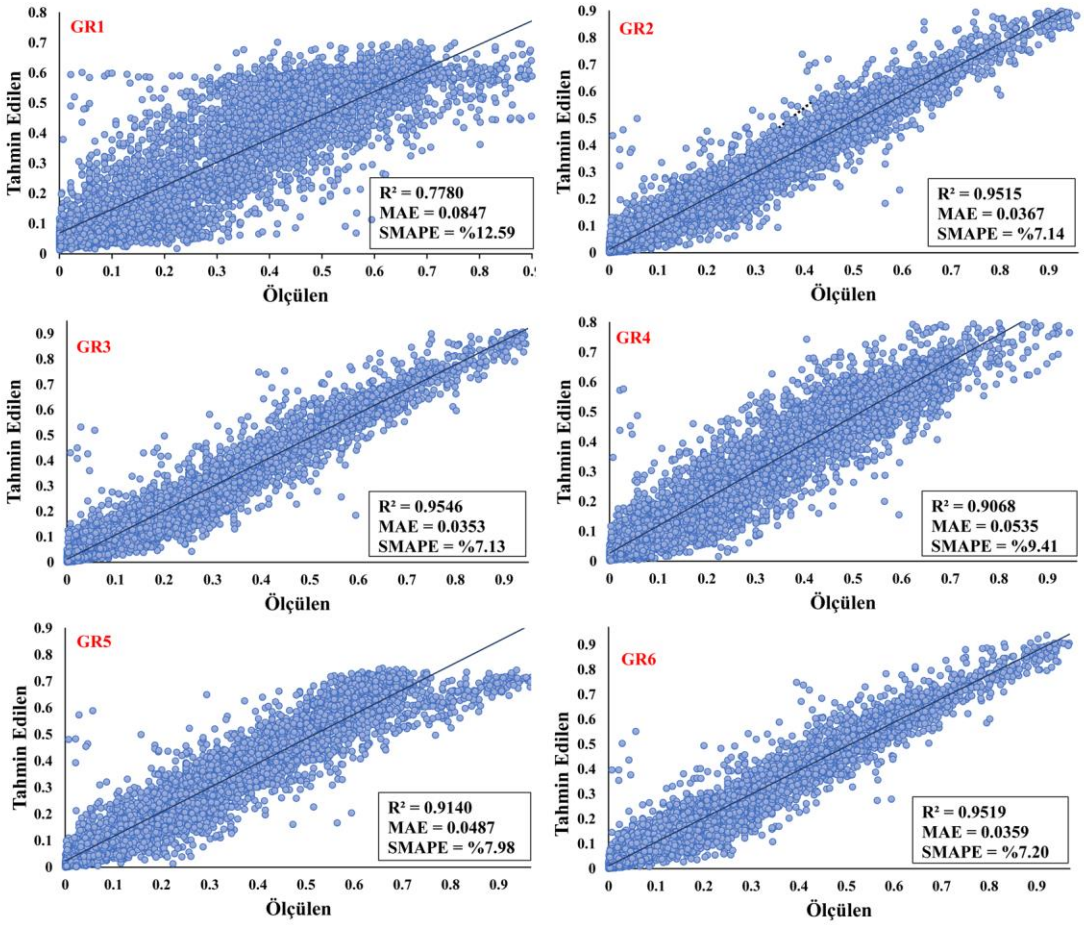


Şekil 5.6. Antalya ili için geliştirilen en iyi modelin ÇKİBSA performans grafiği

Antalya ilinin test verilerine göre ölçülen ve tahmin edilen değerleri arasındaki hata kutu grafikleri Şekil 5.7'de verilmiştir. Ortalama hata seviyelerine göre GR2, GR3 ve GR6 veri grupları ile geliştirilen tahmin modelleri en az hataya sahip modeller olarak hesaplanmıştır. GR1-GR6 veri grubu kullanılarak hesaplanan en başarılı modellerin ölçülen ve tahmin edilen değerleri arasındaki dağılım grafikleri Şekil 5.8'te verilmiştir.



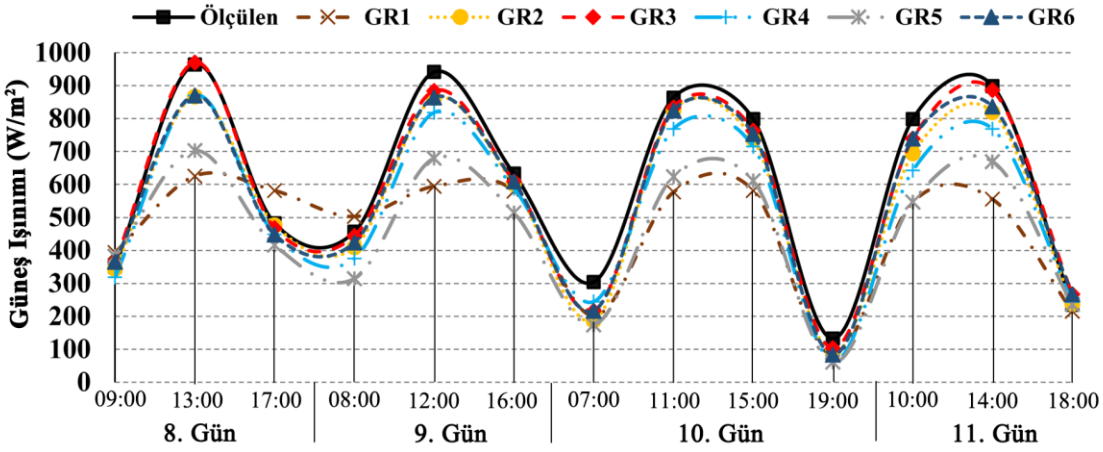
Şekil 5.7. Antalya ili veri gruplarına göre ölçülen ve tahmin edilen değerler arasındaki hata kutu grafiği



Şekil 5.8. Antalya ili test veri gruplarına göre geliştirilen ÇKİBSA modellerinin KGI tahmin dağılım grafikleri

Dağılım grafiklerinden, GR2, GR3 ve GR6 veri grupları için ortalama **%95** ile yüksek doğrulukta bir tahmin seviyesine ulaşılırken, GR4 ve GR5 için ortalama **%91** ile iyi düzeyde bir tahmin performansı elde edildiği anlaşılmaktadır. Girdi veri grupları arasındaki en kötü tahmin sonucuna GR1 veri gurubu ile geliştirilen tahmin modelinde ulaşılmıştır. Antalya ili için VM süreçleri sonunda CfsSE özellik seçim fonksiyonu kullanılarak oluşturulan GR1 veri grubuna ait özelliklerin KGI tahmin etmede yetersiz kaldığı sonucuna varılabilir. En başarılı veri grubu modellerinden az da olsa GR3 tabanlı tahmin modelinin KGI'nı daha iyi tahmin ettiği anlaşılmıştır. Antalya ilinin girdi veri gruplarına göre geliştirilen en iyi modellerinin güneş ışınım tahminleri ile gerçek ölçüm değerlerinin karşılaştırmalı performans grafiği Şekil 5.9'da verilmiştir. Test girdi verilerinden rastgele seçilen 2006 yılının Temmuz ayının 8-11 günleri arası dört günlük saatlik verileri kullanılmıştır. Her bir gün için rastgele seçilerek oluşturulan test

verilerinin ilgili saat dilimlerinde ölçüm ve tahmin sonuçları karşılaştırılmasında, VM süreci destekli veri gruplarının genel olarak güneş ışınımını tahmin etmede oldukça iyi bir performans gösterdiği söylenebilir.



Şekil 5.9. Antalya ili 2006 yılı Temmuz ayı dört günlük ölçülen ve tahmin edilen saatlik verilerin karşılaştırmalı performans grafiği

5.1.3. Burdur ilinin tahmin bulguları

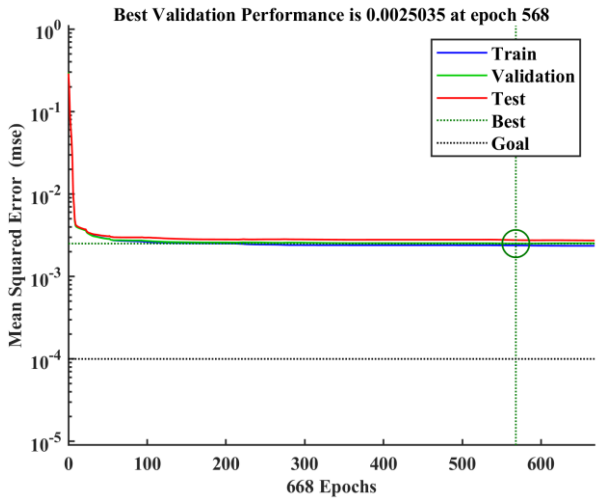
Akdeniz bölgesinin batı iç kesiminde bulunan Burdur ili 29º-30º doğu boylamları ve 36º-37º kuzey enlemleri arasında yer almaktadır. İlin yüzölçümü 7,176 Km² ve deniz seviyesinden yüksekliği ortalama 1,000 metreyi bulmaktadır. Akdeniz iklimi ile karasal iklim arasında bir geçiş iklimi görüldüğünden kışın sıcaklıklar -16.7 °C düşmekte, yazın 41 °C kadar çıkmaktadır. Yıllık sıcaklık ortalaması 13.3 °C, ortalama güneşlenme süresi 7.5 saat seviyelerinde hesaplanmıştır (MGM, 2020). Burdur ilinin meteorolojik verileri 17,238 numaralı istasyondan alınmıştır. Seçilen istasyonun enlemi 37.722, boylamı 30.294 ve deniz seviyesinden yüksekliği 957 m'dir.

Burdur ili için seçilen GR1-GR6 arası girdi verileri kullanılarak ÇKİBSA tahmin modelleri geliştirilmiştir. Seçim grupları temelinde oluşturulan her bir sinir ağının tahmin performansları istatistiksel olarak değerlendirildiğinde hesaplanan en başarılı modellerin tahmin sonuçları Çizelge 5.3'de gösterilmiştir. ClassAE özellik seçim fonksiyonu temelli belirlenen 7 değişkenli GR2 veri grubu, CSE-WSE özellik seçim fonksiyonuna temelli belirlenen 5 özellikli GR3 veri grubu, RAE özellik seçim fonksiyonuna temelli belirlenen 4 özellikli GR5 veri grubu ve GR6 veri grubunda tüm değişkenlerin kullanıldığı modellerin eğitim tahmin performansları R istatistik ölçeğine

göre sırasıyla **0.9799, 0.9835, 0.9786** ve **0.9819** değerleri ile birbirlerine oldukça yakın bulunmuştur. GR2, GR3, GR5 ve GR6 veri gruplarının test tahmin performansları ölçülen gerçek değerlerle karşılaştırıldığında SMAPE ölçeği için sırasıyla **7.69, 7.49, 7.74** ve **7.57** değerleri bulunurken, MAE ölçeğine göre sırasıyla **0.0394, 0.0354, 0.0397** ve **0.0360** değerleri hesaplanmıştır. GR3 veri grubuyla geliştirilen modelin diğer veri gruplarına nazaran $\mathbf{R^2}$ **= %97** değeriyle Burdur ilinin KGI'nı daha iyi bir şekilde modellediği sonucuna varılmıştır. İlgili veri grubu için tasarlanan 49 gizli katman nöronlu ağ eğitim modelinin 1. tekrar sonucundaki ağ performans grafiği Şekil 5.7'de gösterilmiştir. En başarılı performansa 568 epok sonucunda ulaşılmıştır. MSE tahmin performansı yaklaşık **0.0025** seviyelerinde hesaplanmıştır.

Çizelge 5.3. Burdur ili veri gruplarına göre geliştirilen en başarılı modellerin eğitim ve test tahmin sonuçları

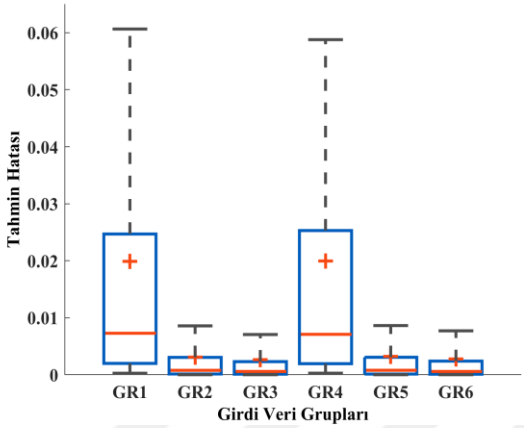
		Modellerin Eğitim Performansı						Modellerin Test Performansı			
Veri Grubu	**Gizli Katman Nöron Sayısı**	**Eğitim Tekrar Sayısı**	**Eğitim Verisi MSE**	**Test Verisi MSE**	**Doğrulama Verisi MSE**	**Model MSE**	**Model R**	**RMSE**	**MAE**	**SMAPE (%)**	**R^2**
GR1	30	4	0.0199	0.0198	0.0204	0.0205	0.8534	0.1411	0.1102	14.33	0.7363
GR2	44	1	0.0028	0.0032	0.0030	0.0031	0.9799	0.0558	0.0394	7.69	0.9588
GR3	***49***	***1***	***0.0023***	***0.0028***	***0.0025***	***0.0025***	***0.9835***	***0.0514***	***0.0354***	***7.49***	***0.9650***
GR4	11	2	0.0202	0.0204	0.0206	0.0207	0.8523	0.1412	0.1102	14.35	0.7361
GR5	31	5	0.0031	0.0033	0.0032	0.0032	0.9786	0.0567	0.0397	7.74	0.9574
GR6	45	1	0.0026	0.0027	0.0027	0.0027	0.9819	0.0527	0.0360	7.57	0.9632



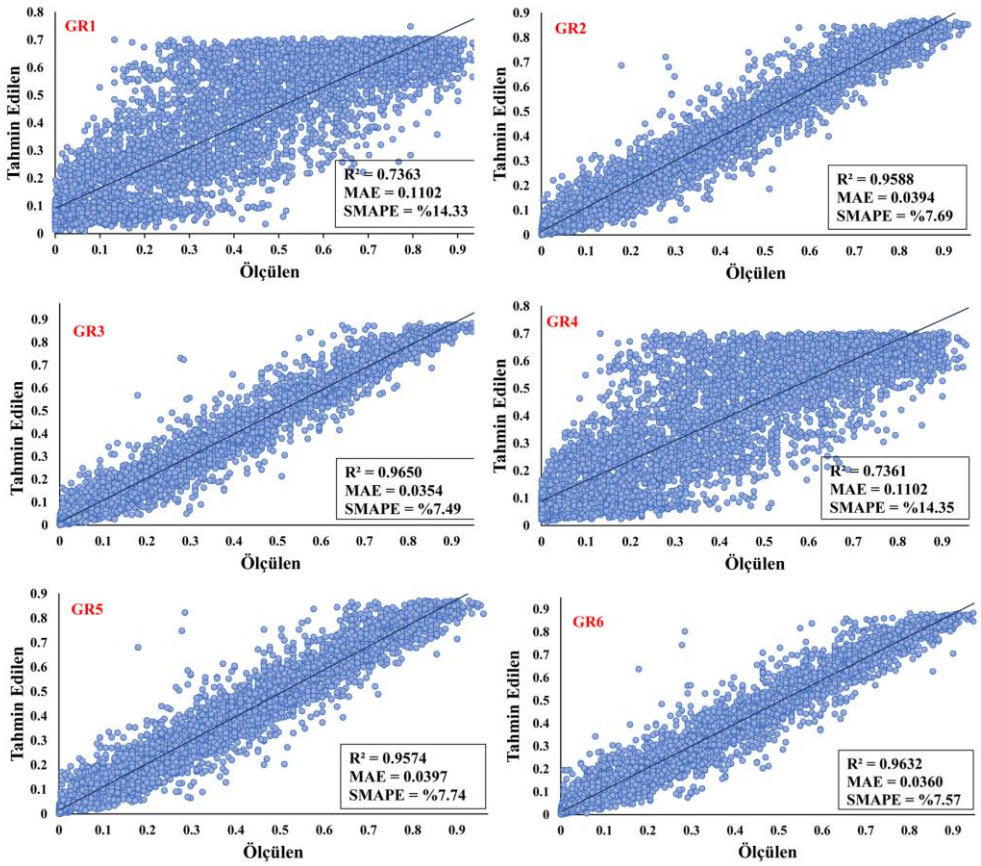
Şekil 5.10. Burdur ili için geliştirilen en iyi modelin ÇKİBSA performans grafiği

Burdur ilinin test verilerine göre ölçülen ve tahmin edilen değerleri arasındaki hata kutu grafikleri Şekil 5.10'da verilmiştir. Ortalama hata seviyelerine göre GR2,

GR3, GR5 ve GR6 veri grupları ile geliştirilen tahmin modelleri en az hataya sahip modeller olarak görülmektedir. GR1-GR6 veri grubu kullanılarak hesaplanan en başarılı modellerin ölçülen ve tahmin edilen değerleri arasındaki dağılım grafikleri Şekil 5.11'de verilmiştir.

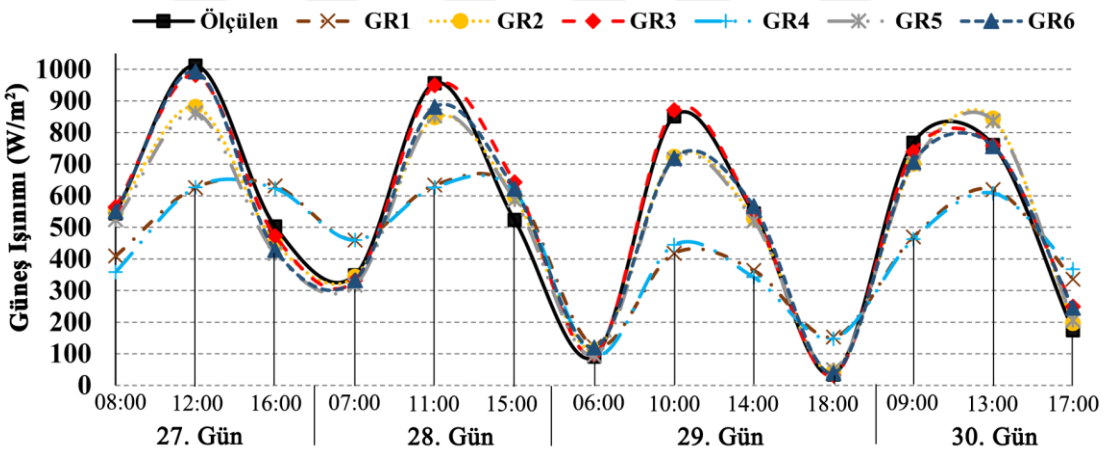


Şekil 5.11. Burdur ili veri gruplarına göre ölçülen ve tahmin edilen değerler arasındaki hata kutu grafiği



Şekil 5.12. Burdur ili test veri gruplarına göre geliştirilen ÇKİBSA modellerinin KGI tahmin dağılım grafikleri

Yukarıda verilen dağılım grafiklerinden GR2, GR3, GR5 ve GR6 özellik gruplarının hedef ilin KGI'nı tahmin etmede başarılı bir performans seviyesine ulaştığı görülmüştür. Aynı grafiklerden GR1 ve GR4 girdi veri grupları ile geliştirilen tahmin modellerinin güneş ışınımı tahmin etmede yetersiz kaldığı sonucuna varılmıştır. Tahmin sonuçlarından, CfsSE ve CorrAE özellik seçim fonksiyonlarının hedef ilin optimal verilerinin seçiminde başarısız olduğu, diğer dört farklı özellik seçim yönteminin başarı bir şekilde en uygun girdi özelliklerini hesapladığı anlaşılmaktadır. VM süreci destekli özellik seçim girdi veri grupları, tüm girdi değişkenleri kullanılarak hesaplanan tahmin performansları ile karşılaştırıldıklarında, ÇKİBSA modellerinde oldukça başarılı sonuçlar verdiği gözlemlenmiştir. Burdur ilinin GR1-GR6 arası girdi veri grupları için güneş ışınım tahminleri ile gerçek ölçüm sonuçlarının karşılaştırmalı performans grafiği Şekil 5.12'de verilmiştir. Test girdi verilerinden rastgele seçilen 2002 yılının Nisan ayının 27-30 günleri arası dört günlük saatlik verileri kullanılmıştır. Grafikten, GR3 girdi veri grubu ile geliştirilen ÇKİBSA modelinin hedef ilin KGI'nı başarılı bir şekilde tahmin ettiği açık bir şekilde görülmektedir.



Şekil 5.13. Burdur ili 2002 yılı Nisan ayı dört günlük ölçülen ve tahmin edilen saatlik verilerin karşılaştırmalı performans grafiği

5.1.4. Hatay ilinin tahmin bulguları

Hatay, Akdeniz bölgesinin doğusunda 35º-37º doğu boylamları ve 35º-36º kuzey enlemleri arasında yer almaktadır. İlin yüzölçümü 5,403 Km² ve deniz seviyesinden yükseltisi ortalama 85 metreyi bulmaktadır. Yıllık sıcaklık ortalaması 18.3 °C, ortalama güneşlenme süresi 7.2 saat seviyelerinde hesaplanmıştır (MGM, 2020). Hatay ilinin

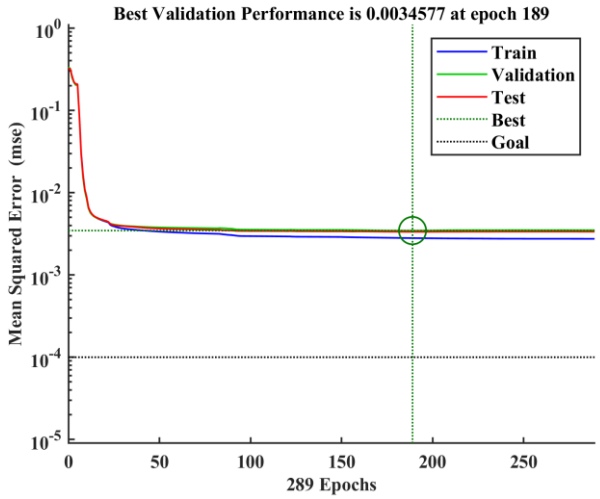
meteorolojik verileri 17,372 numaralı istasyondan alınmıştır. Seçilen istasyonun enlemi 36.205, boylamı 36.151 ve deniz seviyesinden yüksekliği 104 m'dir.

VM süreçleri sonunda oluşturulan GR1-GR7 arası girdi verileri kullanılarak geliştirilen tahmin modelleri istatistiki olarak değerlendirildiğinde hesaplanan en başarılı modellerin tahmin sonuçları Çizelge 5.4'de gösterilmiştir. GR2 veri grubunda seçilen 7, GR3 veri grubunda seçilen 7, GR5 veri grubunda seçilen 4, GR6 veri grubunda seçilen 3 ve GR7 veri grubunda tüm özellikler kullanılarak geliştirilen ÇKİBSA modellerinin tahmin sonuçları birbirlerine oldukça yakın hesaplanmıştır. R istatistik ölçeğine GR2, GR3, GR5, GR6 ve GR7 modellerinin eğitim performansları sırasıyla **0.9762, 0.9799, 0.9752, 0.9740** ve **0.9804** bulunmuştur. Aynı veri grupları için SMAPE ölçeğine göre test tahmin performansları sırasıyla **7.67, 7.63, 7.63, 7.41** ve **7.50** değerleri ile birbirlerine çok yakın bulunurken, MAE ölçeğine göre sırasıyla **0.0424, 0.0399, 0.0434, 0.0432** ve **0.0392** değerleri hesaplanmıştır. Tüm veri grupları içinden GR3 ve GR7 girdi veri grupları ile geliştirilen modellerin tahmin performansları $\mathbf{R^2}$ istatistik ölçeğine göre sırasıyla **0.9548** ve **0.9565** hesaplanarak, Hatay ilinin KGI'nı başarılı bir şekilde tahmin ettiği sonucuna varılmıştır.

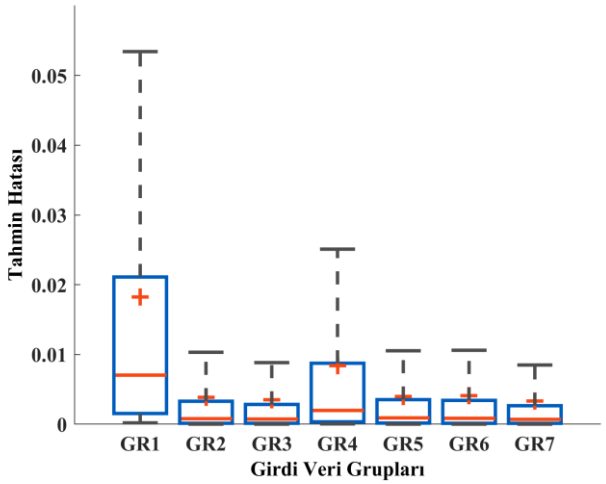
Çizelge 5.4. Hatay ili veri gruplarına göre geliştirilen en başarılı modellerin eğitim ve test tahmin sonuçları

		Modellerin Eğitim Performansı						Modellerin Test Performansı			
Veri Grubu	**Gizli Katman Nöron Sayısı**	**Eğitim Tekrar Sayısı**	**Eğitim Verisi MSE**	**Test Verisi MSE**	**Doğrulama Verisi MSE**	**Model MSE**	**Model R**	**RMSE**	**MAE**	**SMAPE (%)**	**R^2**
GR1	19	3	0.0176	0.0175	0.0178	0.0179	0.8769	0.1351	0.1044	13.12	0.7624
GR2	28	2	0.0034	0.0039	0.0036	0.0035	0.9762	0.0622	0.0424	7.67	0.9496
GR3	49	4	0.0028	0.0033	0.0031	0.0035	0.9799	0.0590	0.0399	7.63	0.9548
GR4	45	5	0.0077	0.0082	0.0078	0.0082	0.9466	0.0917	0.0657	9.72	0.8906
GR5	48	4	0.0036	0.0039	0.0036	0.0038	0.9752	0.0630	0.0434	7.63	0.9483
GR6	27	1	0.0038	0.0040	0.0039	0.0044	0.9740	0.0638	0.0432	7.41	0.9470
GR7	***48***	***2***	***0.0028***	***0.0032***	***0.0030***	***0.0035***	***0.9804***	***0.0579***	***0.0392***	***7.50***	***0.9565***

Hatay ilinin KGI başarılı bir şekilde modelleyen GR3 ve GR7 grupları arasında az da olsa daha iyi performans gösteren GR7 veri grubunun 48 gizli katman nöronlu ağ eğitim modelinin 2. tekrar sonucundaki ağ performans grafiği Şekil 5.14'de verilmiştir. En başarılı performansa 189 epok sonucunda ulaşılmıştır. MSE tahmin performansı yaklaşık **0.0035** seviyelerinde hesaplanmıştır. Hatay ilinin tüm girdi veri gruplarına göre test tahmin performanslarının ölçülen ve tahmin edilen değerleri arasındaki hata kutu grafiği Şekil 5.15'te gösterilmiştir.

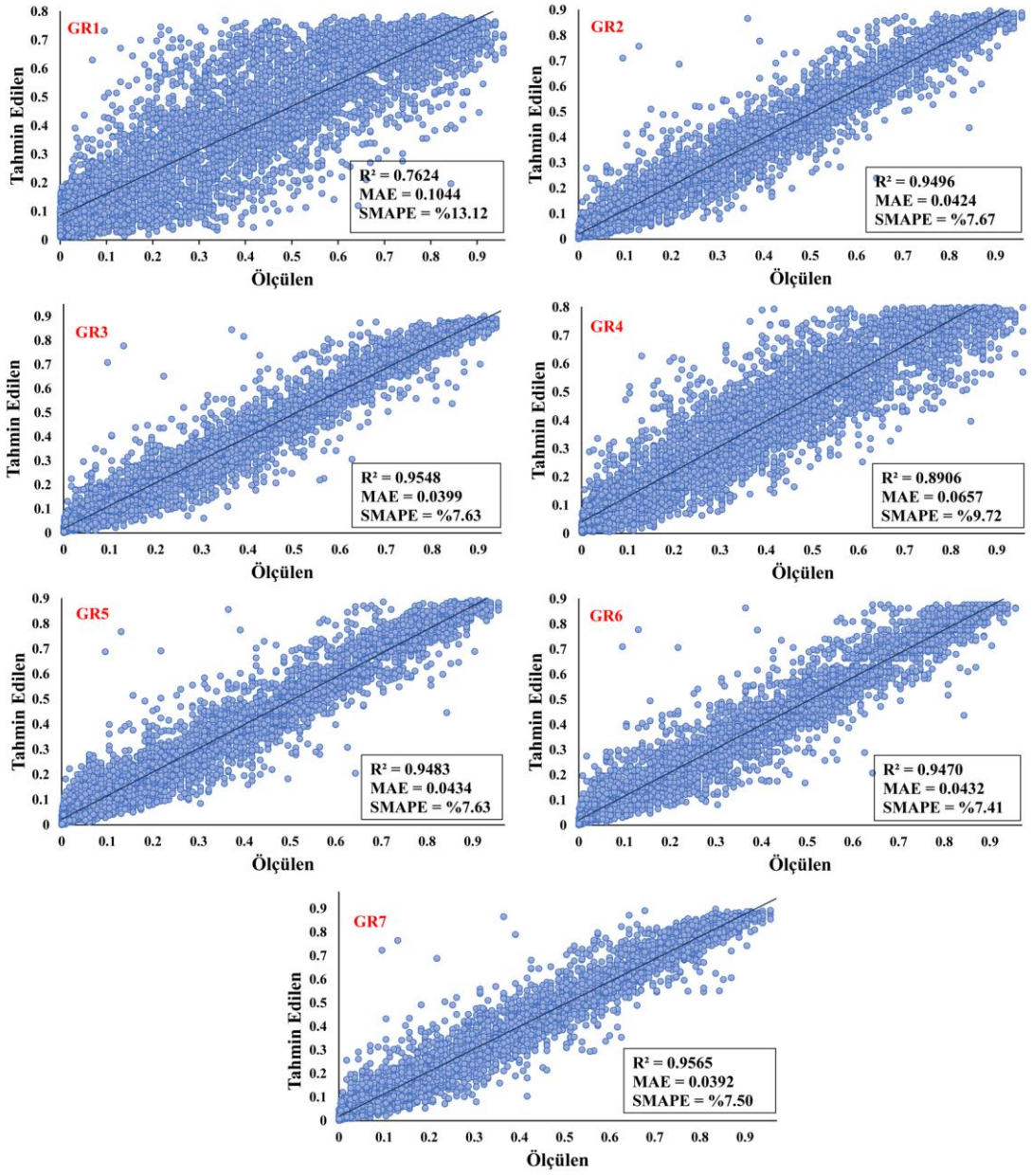


Şekil 5.14. Hatay ili için geliştirilen en iyi modelin ÇKİBSA performans grafiği



Şekil 5.15. Hatay ili veri gruplarına göre ölçülen ve tahmin edilen değerler arasındaki hata kutu grafiği

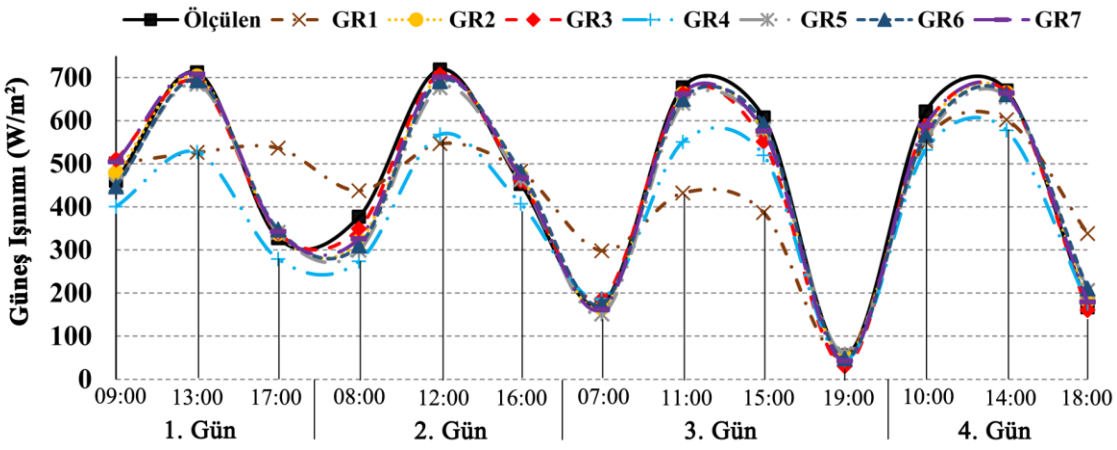
Ortalama hata seviyelerine göre GR2, GR3, GR5, GR6 ve GR7 veri grupları ile geliştirilen tahmin modelleri en az hataya sahip modeller olarak hesaplanmışlardır. Bu gruplar içinde tahmin hatası daha az olan GR3 ve GR7 veri gruplarının birbirlerine benzeştiği görülmektedir. GR1-GR7 veri grubu kullanılarak hesaplanan en başarılı modellerin ölçülen ve tahmin edilen değerleri arasındaki dağılım grafikleri Şekil 5.16'da verilmiştir.



Şekil 5.16. Hatay ili test veri gruplarına göre geliştirilen ÇKİBSA modellerinin KGI tahmin dağılım grafikleri

Dağılım grafiklerinden, GR2, GR3, GR5, GR6 ve GR7 veri grupları ile geliştirilen modellerin tahmin performanslarının birbirlerine benzeştiği görülmektedir. Hatay ili için ClassAE, CSE, WSE ve RAE özellik seçim fonksiyonu kullanılarak seçilen girdi veri grupları ile geliştirilen modeller KGI'nı daha iyi tahmin ederken, CfsSE ve CorrAE ile geliştirilen modellerin tahmin doğrulukları düşük kalmıştır. Süreç sonunda, VM süreci destekli ÇKİBSA ML modellerinin KGI'nı başarılı bir şekilde tahmin ettiği sonucuna varılabilmektedir. Hatay ilinin GR1-GR7 arası geliştirilen test modellerinin güneş ışınım tahminleri ile gerçek ölçüm değerlerinin karşılaştırmalı

performans grafiği Şekil 5.17'de gösterilmiştir. Test girdi verilerinden rastgele seçilen 2003 yılının Mayıs ayının 1-4'ncü günleri arası dört günlük saatlik verileri kullanılmıştır. GR2, GR3, GR5, GR6 ve GR7 veri grupları test verilerinin ilgili saat dilimlerinde ölçüm ve tahmin sonuçları karşılaştırıldığında genel olarak KGI'nı tahmin etmede oldukça iyi bir performans gösterdiği sonucuna varılmıştır.



Şekil 5.17. Hatay ili 2002 yılı Nisan ayı dört günlük ölçülen ve tahmin edilen saatlik verilerin karşılaştırmalı performans grafiği

5.1.5. Isparta ilinin tahmin bulguları

Isparta ili Akdeniz bölgesinin batısında 30º-31º doğu boylamları ve 37º-38º kuzey enlemleri arasında yer almaktadır. İlin yüzölçümü 8,933 Km² ve deniz seviyesinden yükseltisi ortalama 1050 metreyi bulmaktadır. Akdeniz ikliminden karasal iklime geçiş bölgesinde yer alan ilin kışın ölçülen en düşük sıcaklığı -21 °C ve yazın en yüksek 42.3 °C olarak ölçülmüş, sıcaklık ortalaması 12.3 °C, ortalama güneşlenme süresi 7.1 saat seviyelerinde hesaplanmıştır (MGM, 2020). Isparta ilinin meteorolojik verileri 17,240 numaralı istasyondan alınmıştır. Seçilen istasyonun enlemi 37.785, boylamı 30.568 ve deniz seviyesinden yüksekliği 994 m olarak ölçülmüştür.

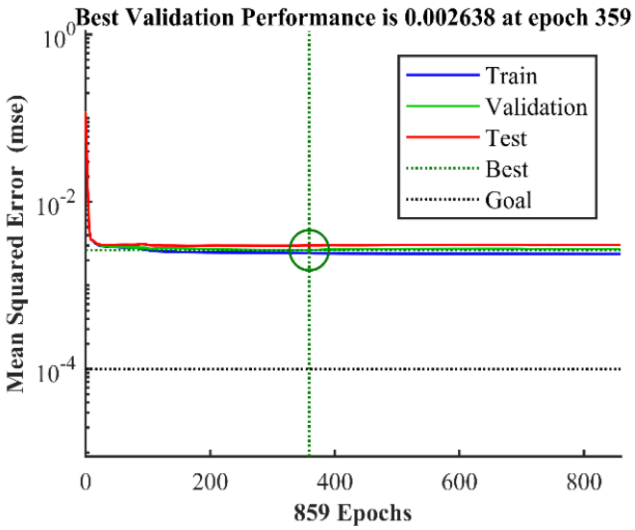
Hedef il için toplanan meteorolojik ve kategorik veri seti VM süreçlerine sokularak GR1-GR6 arası girdi veri grupları oluşturulmuştur. Yazılan program kodu ile geliştirilen ÇKİBSA modelleri yardımıyla KGI'nı tahmin etmek amacıyla her bir veri grubu için 250 farklı sinir ağı modeli arasından en başarılı olanlar Çizelge 5.5'de verilmiştir. GR2, GR3 ve GR6 veri grupları ile geliştirilen en iyi modellerin eğitim performansları R istatistik ölçeğine göre sırasıyla **0.9724, 0.9768** ve **0.9745** bulunarak birbirlerine oldukça yakın yüksek seviyeli bir tahmin korelasyonu hesaplanmıştır. GR5

girdi veri grubu **R = 0.9442** değeri iyi düzeyde bir tahmin performansına ulaşırken, en kötü performans GR1 ve GR4 veri grubunda bulunmuştur. Geliştirilen modellerin test performansları incelendiğinde R^2 ölçeğine göre GR2, GR3 ve GR6 veri gruplarının tahmin performansı sırasıyla **0.9444, 0.9488** ve **0.9477** hesaplanarak birbirlerine oldukça yakın tahmin performansı göstermişlerdir. SMAPE ve MAE ölçeklerine göre sırasıyla **%7.77** ve **0.0352** değerleri ile CSE ve WSE özellik seçim fonksiyonları kullanılarak seçilen 7 girdili GR3 veri grubu tahmin modeli KGI'nı diğer veri gruplarına nazaran daha yüksek doğrulukta tahmin etmiştir. GR3 girdi veri grubu için tasarlanan 48 gizli katman nöronlu ağ eğitim modelinin 1'nci tekrar sonucundaki ağ performans grafiği Şekil 5.18'de gösterilmiştir. En başarılı performansa 359 epok sonucunda ulaşılmıştır. MSE tahmin performansı yaklaşık 0.0026 seviyelerinde hesaplanmıştır.

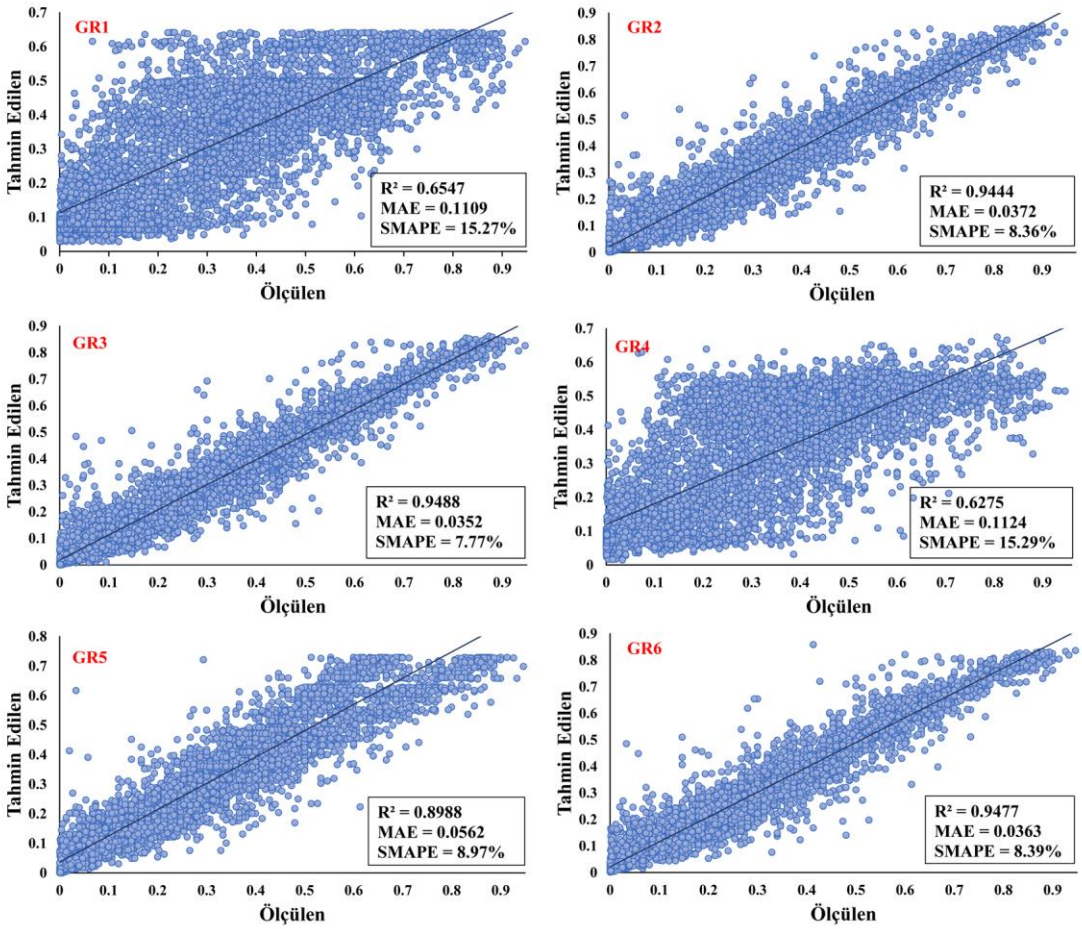
Çizelge 5.5. Isparta ili veri gruplarına göre geliştirilen en başarılı modellerin eğitim ve test tahmin sonuçları

		Modellerin Eğitim Performansı						Modellerin Test Performansı			
Veri Grubu	**Gizli Katman Nöron Sayısı**	**Eğitim Tekrar Sayısı**	**Eğitim Verisi MSE**	**Test Verisi MSE**	**Doğrulama Verisi MSE**	**Model MSE**	**Model R**	**RMSE**	**MAE**	**SMAPE (%)**	**R^2**
GR1	18	2	0.0198	0.0202	0.0202	0.0201	0.7970	0.1392	0.1109	15.27	0.6547
GR2	28	1	0.0029	0.0031	0.0034	0.0030	0.9724	0.0559	0.0372	8.36	0.9444
GR3	***48***	***1***	***0.0024***	***0.0029***	***0.0026***	***0.0026***	***0.9768***	***0.0536***	***0.0352***	***7.77***	***0.9488***
GR4	32	1	0.0212	0.0217	0.0216	0.0216	0.7803	0.1446	0.1124	15.29	0.6275
GR5	21	5	0.0059	0.0057	0.0063	0.0059	0.9442	0.0754	0.0562	8.97	0.8988
GR6	25	3	0.0027	0.0030	0.0029	0.0028	0.9745	0.0542	0.0363	8.39	0.9477

Girdi veri gruplarına göre ölçülen tahmin edilen değerler arasındaki dağılım grafikleri Şekil 5.19'da verilmiştir. Bu grafiklerden GR2, GR3 ve GR6 veri grupları ile geliştirilen ÇKİBSA modellerinin KGI'nı başarılı bir şekilde tahmin ettiği görülmektedir. GR1 ve GR4 veri gruplarının seçiminde kullanılan CfsSE ve CorrAE özellik seçim fonksiyonlarının KGI'nın tahmin edilmesinde oldukça düşük bir performans seviyesinde kaldığı anlaşılmaktadır.



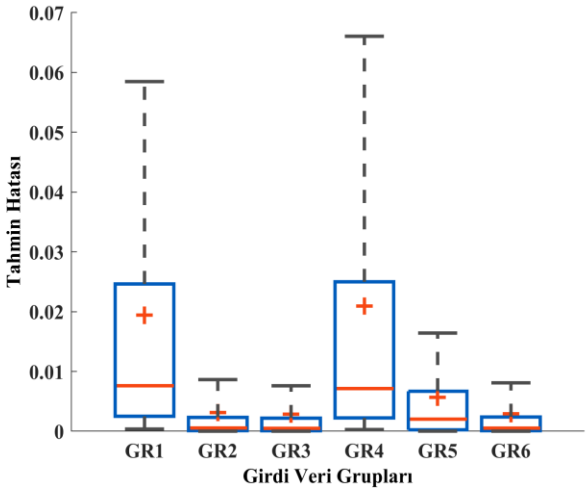
Şekil 5.18. Isparta ili için geliştirilen en iyi modelin ÇKİBSA performans grafiği



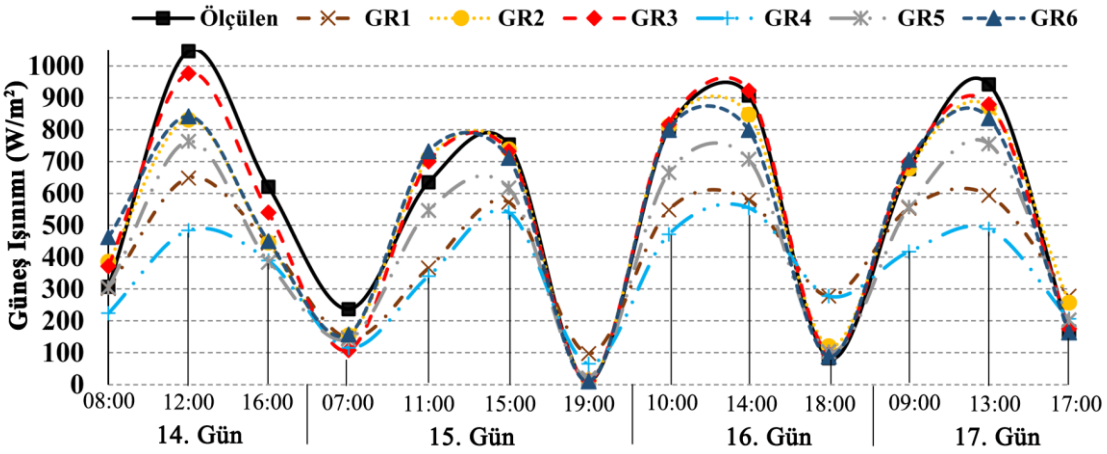
Şekil 5.19. Isparta ili test veri gruplarına göre geliştirilen ÇKİBSA modellerinin KGI tahmin dağılım grafikleri

Isparta ilinin GR1-GR6 veri gruplarına göre ölçülen ve tahmin edilen değerleri arasındaki hata kutu grafiği Şekil 5.20'de verilmiştir. Grafikten GR2, GR3 ve GR6 veri

grupları ile geliştirilen modellerin en az hataya sahip tahminler olduğu açık bir biçimde görülmektedir. GR3 veri grubu modelinin, girdi veri grupları içinde küçükte olsa daha az tahmin hatasına sahip olduğu anlaşılmaktadır. GR1-GR6 veri grubu kullanılarak hesaplanan en başarılı modellerin ölçülen ve tahmin edilen değerleri arasındaki karşılaştırmalı performans grafiği Şekil 5.21'de verilmiştir. Test girdi verilerinden rastgele seçilen 2002 yılının Mayıs ayının 14-17 günleri arası dört günlük saatlik verileri kullanılmıştır. GR2, GR3 ve GR6 veri gruplarının ölçüm ve tahmin sonuçları karşılaştırıldığında genel olarak KGI'nı tahmin etmede oldukça iyi bir performans gösterdiği ve GR3 veri grubu kullanılarak geliştirilen modelin daha kesin tahmin doğruluğuna ulaştığı sonucuna varılmıştır.



Şekil 5.20. Isparta ili veri gruplarına göre ölçülen ve tahmin edilen değerler arasındaki hata kutu grafiği



Şekil 5.21. Isparta ili 2002 yılı Mayıs ayı dört günlük ölçülen ve tahmin edilen saatlik verilerin karşılaştırmalı performans grafiği

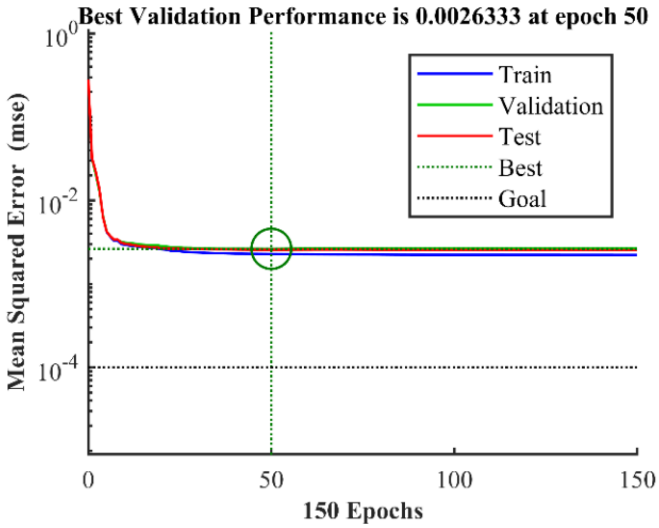
5.1.6. Kahramanmaraş ilinin tahmin bulguları

Akdeniz bölgesinin doğusunda yer alan Kahramanmaraş ilinin yüzölçümü 14,346 km^2 ve deniz seviyesinden yüksekliği ortalama 568 metredir. Hedef il, 36º-37º doğu boylamları ile 37º-38º kuzey enlemleri arasında yer almaktadır. Coğrafi konumunun etkisi ile farklı iklim tiplerinin etkisinde kalsa da Akdeniz ikliminin karakteristik özelliklerine yakın bir iklim özelliği göstermektedir. Yıllık sıcaklık ortalaması 16.7 iken yaz aylarında sıcaklıklar 45 derecelere kadar çıkabilmekte, kış aylarında -9.6 derecelere kadar düşebilmektedir. Hedef ilin ortalama güneşlenme süresi 6.8 saat olarak hesaplanmıştır (MGM, 2020). Kahramanmaraş ilinin meteorolojik verileri 17,255 numaralı istasyondan ölçülmüştür. Seçilen istasyonun enlemi 37.576, boylamı 36.915 ve deniz seviyesinden yüksekliği 572 m ölçülmüştür.

Kahramanmaraş ili GR1-GR6 arası girdi veri grupları kullanılarak geliştirilen ÇKİBSA algoritması temelli tahmin modellerinden KGI'nı en iyi tahmin edenler Çizelge 5.6'da verilmiştir. Geliştirilen en iyi modellerin eğitim performansları R ölçeğine göre istatistiksel olarak değerlendirildiğinde GR2, GR3, GR5 ve GR6 girdi veri grupları için sırasıyla **0.9885**, **0.9800**, **0.9812** ve **0.9805** değerleri hesaplanarak birbirlerine yakın sonuçlara ulaşılmıştır. Test performansları değerlendirildiğinde GR3, GR5 ve GR6 veri gruplarında MAE ölçeğine göre sırasıyla **0.0385, 0.0386** ve **0.0383** değerleri hesaplanarak tahmin sonuçlarının benzeştiği görülmektedir. En iyi test performansına **%97** tahmin doğruluğu ve **%7.79** SMAPE ölçeği ile GR2 veri grubu tahmin modelinde ulaşılmıştır. 40 gizli katman nöronu kullanılarak 2'nci eğitim tekrarında geliştirilen modelin ağ performans grafiği Şekil 5.22'de gösterilmiştir. En başarılı performansa 50 epok sonucunda yaklaşık **0.0026** MSE değerinde ulaşılmıştır.

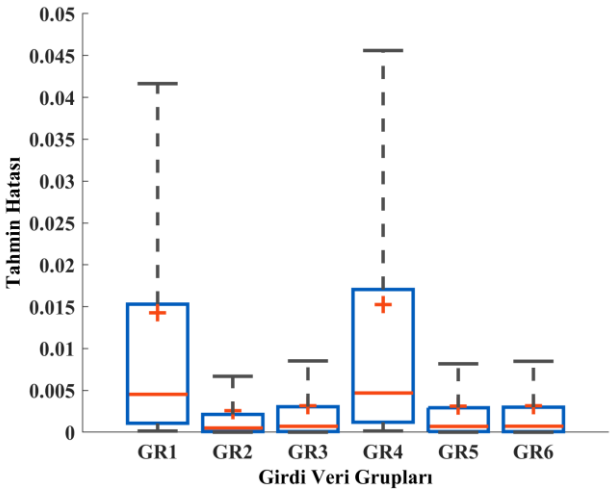
Çizelge 5.5.6. Kahramanmaraş ili veri gruplarına göre geliştirilen en başarılı modellerin eğitim ve test tahmin sonuçları

		Modellerin Eğitim Performansı						**Modellerin Test Performansı**			
Veri Grubu	**Gizli Katman Nöron Sayısı**	**Eğitim Tekrar Sayısı**	**Eğitim Verisi MSE**	**Test Verisi MSE**	**Doğrulama Verisi MSE**	**Model MSE**	**Model R**	**RMSE**	**MAE**	**SMAPE (%)**	**R^2**
GR1	30	3	0.0134	0.0137	0.0141	0.0137	0.9049	0.1181	0.0897	14.25	0.8138
GR2	***40***	***2***	***0.0021***	***0.0026***	***0.0026***	***0.0027***	***0.9845***	***0.0508***	***0.0341***	***7.79***	***0.9656***
GR3	27	1	0.0029	0.0031	0.0028	0.0030	0.9800	0.0556	0.0385	8.23	0.9587
GR4	25	5	0.0145	0.0145	0.015	0.0149	0.8959	0.1234	0.0928	14.37	0.7969
GR5	44	2	0.0027	0.0030	0.0033	0.0028	0.9812	0.0558	0.0386	8.29	0.9585
GR6	39	3	0.0028	0.0028	0.0029	0.0029	0.9805	0.0551	0.0383	8.11	0.9595



Şekil 5.22. Kahramanmaraş ili için geliştirilen en iyi modelin ÇKİBSA performans grafiği

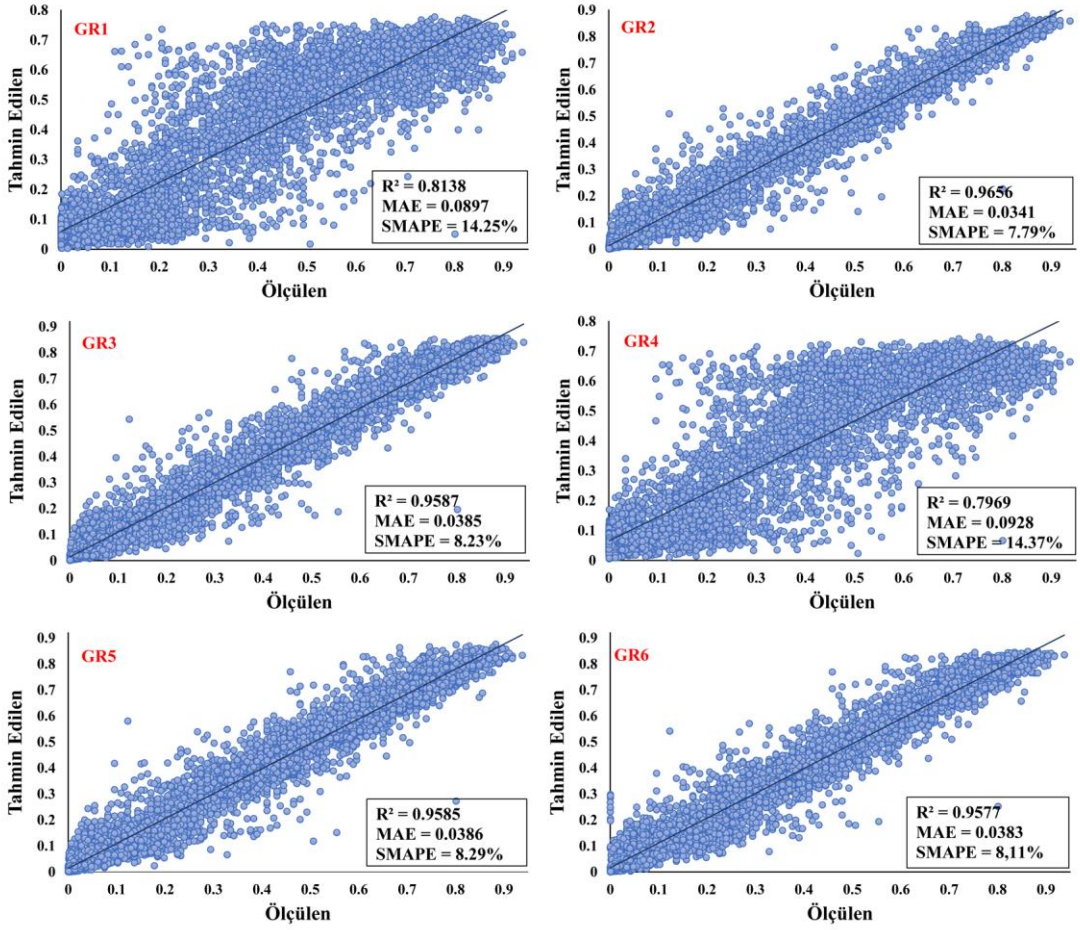
Hedef ilin GR1-GR6 veri gruplarına göre ölçülen ve tahmin edilen değerleri arasındaki hata kutu grafiği Şekil 5.23'te verilmiştir. Grafikten GR3, GR5 ve GR6 veri grupları ile geliştirilen modellerinin hata oranlarının benzeştiği, GR2 veri grubu ile geliştirilen modelin en az hataya sahip tahmin modeli olduğu açık bir biçimde anlaşılmaktadır. GR1-GR6 veri grubu kullanılarak hesaplanan en başarılı modellerin ölçülen ve tahmin edilen değerleri arasındaki dağılım grafikleri Şekil 5.24'te verilmiştir.



Şekil 5.23. Kahramanmaraş ili veri gruplarına göre ölçülen ve tahmin edilen değerler arasındaki hata kutu grafiği

Dağılım grafiklerinden, girdi veri gruplarına göre GR2, GR3, GR5 ve GR6 veri grubu tahmin modellerinin KGI'nı başarılı bir şekilde tahmin ettiği görülmektedir. GR1 ve GR4 veri gruplarının KGI'nın tahmin edilmesinde oldukça düşük bir performans

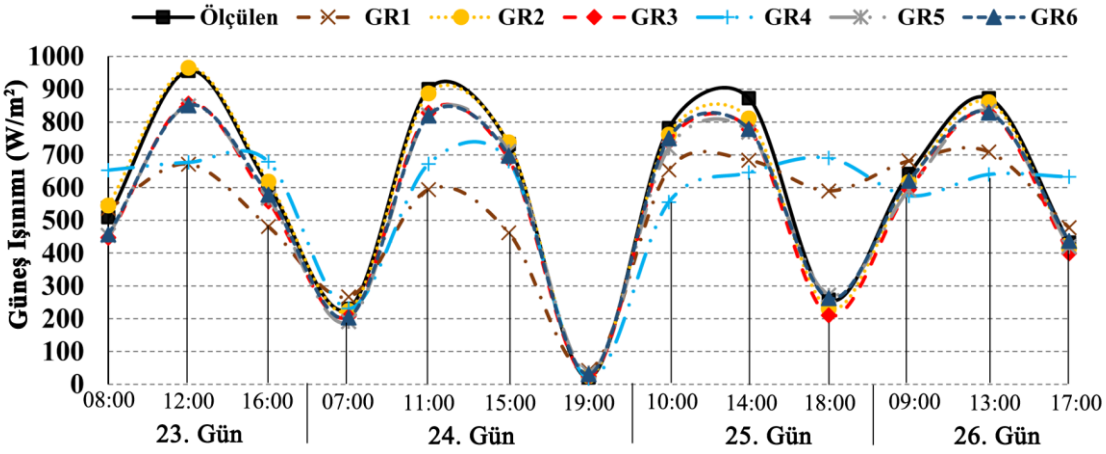
seviyesine ulaşıldığı anlaşılmaktadır. İlgili veri gruplarındaki değişkenlerin seçiminde kullanılan CfsSE ve CorrAE özellik seçim fonksiyonlarının hedef ilin veri özelliklerinin belirlenmesinde başarısız olduğu gözlemlenmiştir. Tahmin modellerindeki en iyi performansa ClassAE özellik seçim fonksiyonu ile belirlenen 5 farklı özellik ile GR2 girdi veri grubunda ulaşılmıştır.



Şekil 5.24. Kahramanmaraş ili test veri gruplarına göre geliştirilen ÇKİBSA modellerinin KGI tahmin dağılım grafikleri

ÇKİBSA algoritmasına göre GR1-GR6 veri grubu kullanılarak hesaplanan en başarılı modellerin ölçülen ve tahmin edilen değerleri arasındaki karşılaştırmalı performans grafiği Şekil 5.25'de verilmiştir. Test girdi verilerinden rastgele seçilen 2003 yılının Haziran ayının 23-26 günleri arası dört günlük saatlik verileri kullanılmıştır. Grafikten GR2 veri grubu kullanılarak geliştirilen modelin KGI'nı tahmin etmede oldukça iyi bir performans gösterdiği anlaşılmakta, GR3, GR5 ve GR6 veri grupları ile geliştirilen modellerin tahmin performanslarının benzeştiği görülmektedir. VM süreçleri kullanılarak seçilen GR2 veri grubu özelliklerinin hedef

ilin güneş ışınımını modellemede yüksek bir tahmin başarısı sağladığı sonucuna varılmıştır.



Şekil 5.25. Kahramanmaraş ili 2003 yılı Haziran ayı dört günlük ölçülen ve tahmin edilen saatlik verilerin karşılaştırmalı performans grafiği

5.1.7. Mersin ilinin tahmin bulguları

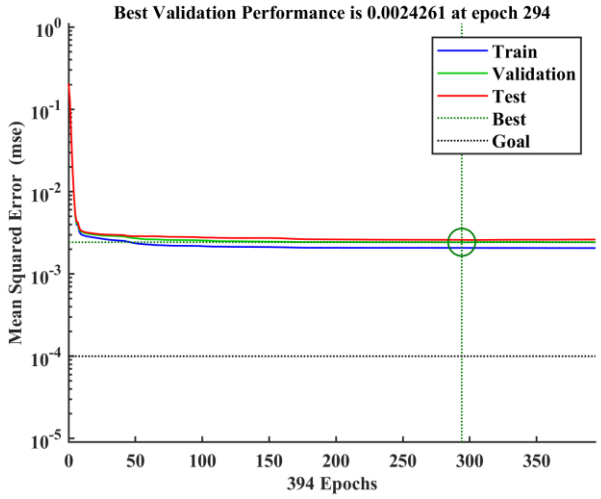
Akdeniz bölgesinin doğusunda bulunan Mersin ili 33°-35° doğu boylamları ile 36°-37° kuzey enlemleri arasında yer almaktadır. İlin ortalama yükseltisi 135 m ve yüzölçümü 15,485 km^2'dir. Sıcaklığın en düşük ölçüldüğü aylar Ocak-Şubat, en yüksek ölçüldüğü aylar ise Temmuz-Ağustos aylarıdır. Sıcaklık ortalaması 19.2 °C, ortalama güneşlenme süresi 7.5 saat ve yıllık güneşli gün sayısı yaklaşık 364 gün olarak hesaplanmıştır (MGM, 2020). Mersin ilinin meteorolojik verileri 17,340 numaralı istasyondan alınmıştır. Seçilen istasyonun enlemi 36.781, boylamı 34.603 ve deniz seviyesinden yüksekliği 7 m olarak ölçülmüştür.

Mersin ili girdi veri gruplarına göre ÇKİBSA algoritması kullanılarak geliştirilen tahmin modellerinden en başarılı olanlar Çizelge 5.7'de verilmiştir. Modellerin eğitiminde GR2, GR3, GR5 ve GR6 veri gruplarının R istatistik ölçeğine göre tahmin performansları sırasıyla **0.9820**, **0.9855**, **0.9818** ve **0.9821** bulunarak birbirlerine oldukça yakın hesaplanmıştır. Geliştirilen modellerin daha önce görmediği test verileri kullanılarak tahmin performansları değerlendirildiğinde SMAPE ölçeğine göre aynı gruplar için sırasıyla **%6.32**, **%6.11**, **%6.13** ve **%6.24** değerleri hesaplanırken, MAE ölçeğine göre sırasıyla **0.0366**, **0.0329**, **0.0365** ve **0.0364** değerleri bulunarak birbirlerine benzer tahmin sonuçlarına ulaşılmıştır. Veri grupları arasında en iyi tahmin

performansına GR3 veri grubu ile **%97** tahmin doğruluğunda ulaşılmıştır. GR3 veri grubu için geliştirilen modelin ağ eğitim performans grafiği Şekil 5.26'da verilmiştir. 40 gizli katman nöronu kullanılarak 3'ncü eğitim tekrarında geliştirilen en başarılı model, 294 epok sonucunda MSE = **0.0024** değerine ulaşmıştır.

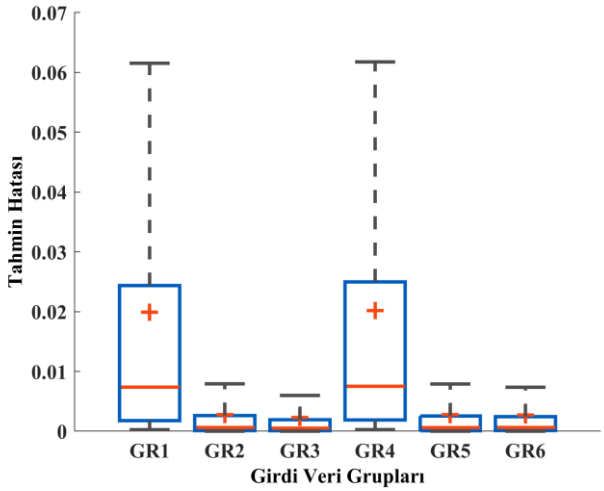
Çizelge 5.7. Mersin ili veri gruplarına göre geliştirilen en başarılı modellerin eğitim ve test tahmin sonuçları

		Modellerin Eğitim Performansı						Modellerin Test Performansı			
Veri Grubu	**Gizli Katman Nöron Sayısı**	**Eğitim Tekrar Sayısı**	**Eğitim Verisi MSE**	**Test Verisi MSE**	**Doğrulama Verisi MSE**	**Model MSE**	**Model R**	**RMSE**	**MAE**	**SMAPE (%)**	**R^2**
GR1	32	2	0.0192	0.0192	0.0196	0.0198	0.8618	0.1410	0.1097	12.70	0.7403
GR2	34	1	0.0027	0.0027	0.0027	0.0026	0.9820	0.0525	0.0366	6.32	0.9641
GR3	***40***	***3***	***0.0021***	***0.0024***	***0.0022***	***0.0024***	***0.9855***	***0.0481***	***0.0329***	***6.11***	***0.9698***
GR4	35	2	0.0200	0.0200	0.0199	0.0201	0.8580	0.1421	0.1106	12.77	0.7364
GR5	30	2	0.0027	0.0028	0.0027	0.0028	0.9818	0.0527	0.0365	6.13	0.9637
GR6	36	4	0.0026	0.0026	0.0027	0.0025	0.9821	0.0522	0.0364	6.24	0.9645

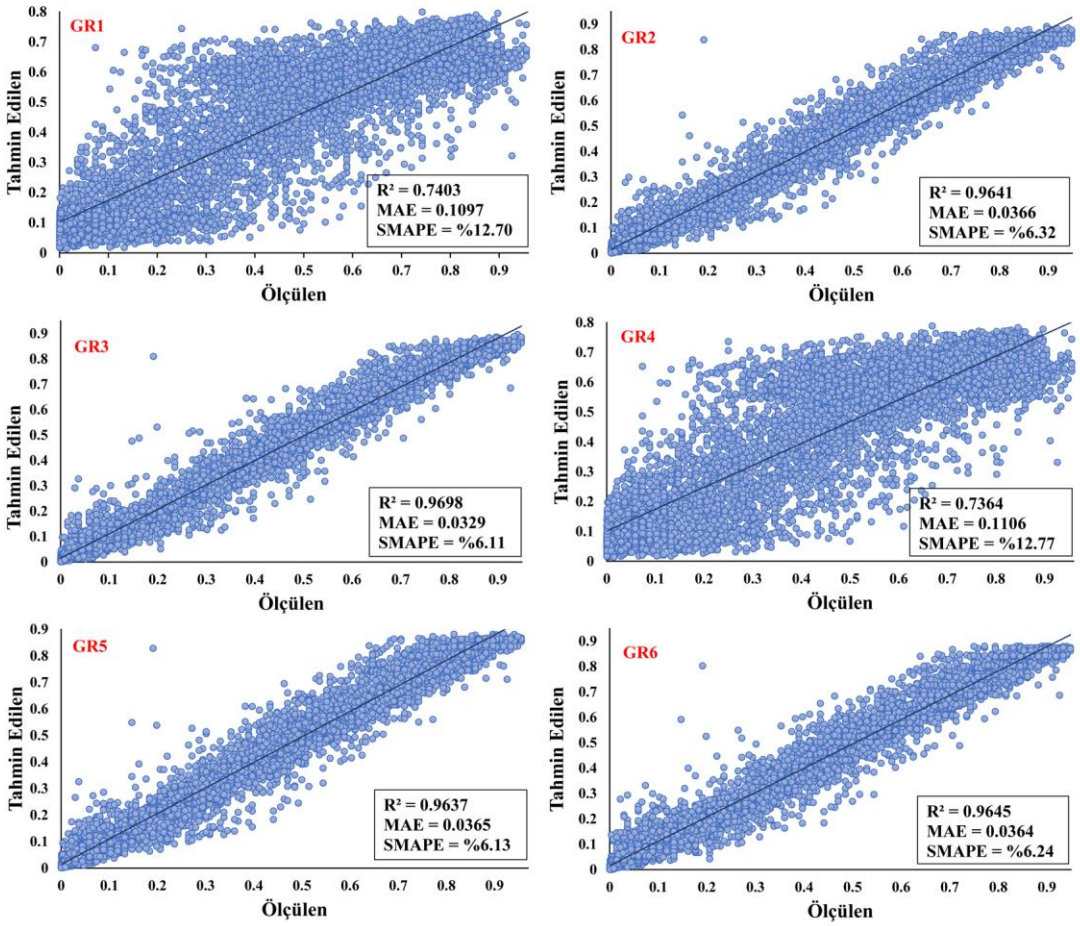


Şekil 5.26. Mersin ili için geliştirilen en iyi modelin ÇKİBSA performans grafiği

Hedef ilin GR1-GR6 veri gruplarına göre ölçülen ve tahmin edilen değerleri arasındaki hata kutu grafiği Şekil 5.27'de verilmiştir. GR2, GR3, GR5 ve GR6 veri grupları ile geliştirilen modellerinin hata oranlarının en az olduğu, GR1 ve GR4 veri gruplarında ise hata oranlarının oldukça arttığı görülmektedir. Bunlar arasında GR3 veri grubu ile geliştirilen modelin en az hataya sahip olduğu anlaşılmaktadır. Tüm girdi veri grupları kullanılarak hesaplanan en başarılı modellerin ölçülen ve tahmin edilen değerleri arasındaki dağılım grafikleri Şekil 5.28'de gösterilmiştir.

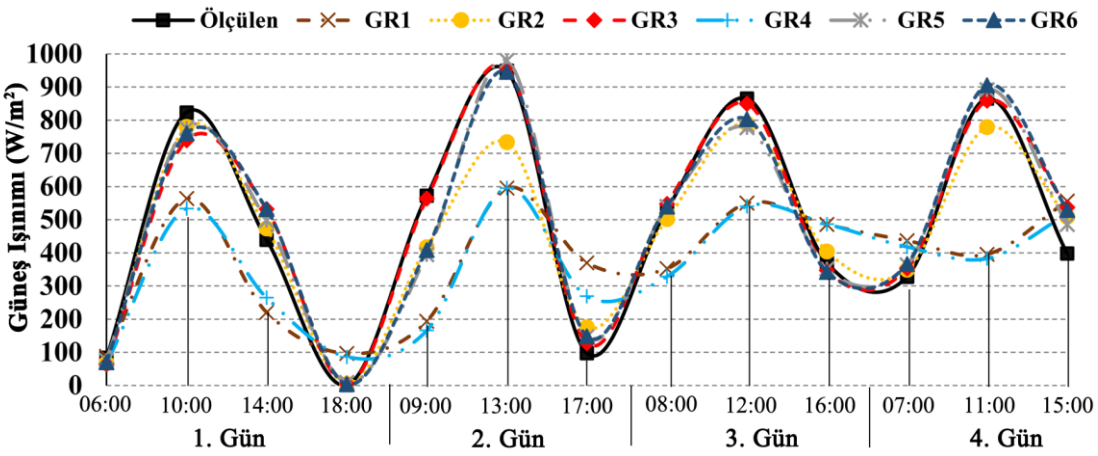


Şekil 5.27. Mersin ili veri gruplarına göre ölçülen ve tahmin edilen değerler arasındaki hata kutu grafiği



Şekil 5.28. Mersin ili test veri gruplarına göre geliştirilen ÇKİBSA modellerinin KGI tahmin dağılım grafikleri

Yukarıda verilen dağılım grafiklerinden VM süreci sonunda oluşturulan GR2, GR3 ve GR5 girdi veri grupları ile geliştirilen tahmin modellerinin KGI'nı başarılı bir şekilde tahmin ettiği görülmekte ve tüm özelliklerin seçildiği GR6 veri grubu ile benzer bir performans seviyesine ulaşıldığı anlaşılmaktadır. Tüm veri grupları içinden GR3 veri grubu tahmin modelinin ölçüm değerlerine oldukça yakın daha iyi bir tahmin korelasyonuna ulaşıldığı sonucuna varılmıştır. GR1-GR6 arası veri gruplarının ölçülen ve tahmin edilen değerleri arasındaki karşılaştırmalı performans grafiği Şekil 5.29'da verilmiştir. Test girdi verilerinden rastgele seçilen 2005 yılının Nisan ayının 1-4 günleri arası dört günlük saatlik verileri kullanılmıştır. GR3, GR5 ve GR6 veri grupları test verilerinin ilgili saat dilimlerinde ölçüm ve tahmin sonuçları karşılaştırmasında genel olarak KGI'nı tahmin etmede oldukça iyi bir performans gösterdiği ve GR3 tahmin sonuçlarının ise diğer veri gruplarına nazaran daha başarılı olduğu anlaşılmaktadır. VM süreçleri sonunda ÇKİBSA algoritması kullanılarak geliştirilen GR3 veri grubu tahmin modelinin Mersin ilinin KGI'nı tahmin etmede kullanılabileceği sonucuna varılmıştır.



Şekil 5.29. Mersin ili 2005 yılı Nisan ayı dört günlük ölçülen ve tahmin edilen saatlik verilerin karşılaştırmalı performans grafiği

5.1.8. Osmaniye ilinin tahmin bulguları

Osmaniye ili, Akdeniz bölgesinin doğusunda 35°-36° doğu boylamları ile 36°-37° kuzey enlemleri arasında yer almaktadır. İlin ortalama yükseltisi 121 m, yüzölçümü 3,280 km^2 ve denize 20 km mesafededir. Osmaniye Akdeniz ikliminin karakteristik özelliklerinin hüküm sürdüğü bir konumda bulunmaktadır. Yıllık en yüksek sıcaklık ortalaması 24.9 °C, en düşük sıcaklık ortalaması 12.7 °C ve ortalama sıcaklık 18.6 °C

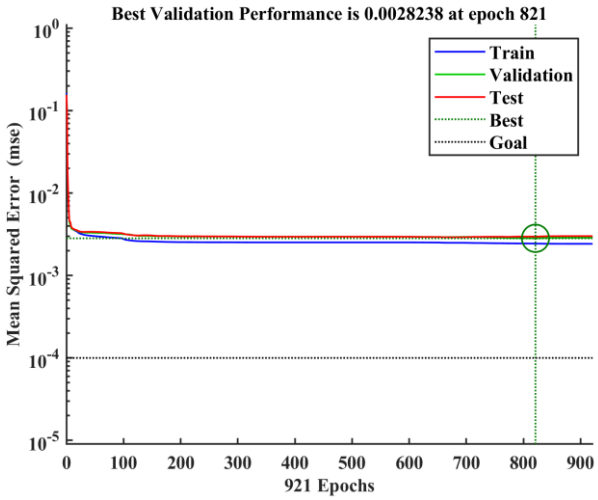
olarak hesaplanmıştır. Ortalama güneşlenme süresi yaklaşık olarak 5.6 saat olarak ölçülmüştür (MGM, 2020). Hedef ilin meteorolojik verileri 17,355 numaralı MGM istasyonu tarafından ölçülmüştür. Seçilen istasyonun enlemi 37.102, boylamı 36.254 ve deniz seviyesinden yüksekliği 94 metredir.

Osmaniye ili meteorolojik ve kategorik girdi verileri VM süreçleri sonunda düzenlenip GR1-GR6 arası oluşturulan girdi verileri yardımıyla geliştirilen ÇKİBSA tahmin modellerinin performans sonuçları Çizelge 5.8'de verilmiştir. En başarılı eğitim performansına 45 gizli katman nöronu ile 5'nci tekrarda GR3 ve 42 gizli katman nöronu ile 3'ncü tekrarda GR6 veri grupları için **MSE = 0.0028** değeri ile ulaşılmıştır. Her iki veri grubu ile geliştirilen en başarılı modellerin R ölçeğine göre eğitim performansı yaklaşık **0.98** seviyelerindedir. Veri gruplarının test tahmin performansları değerlendirildiğinde **MAE = 0.0372** ve **SMAPE = %6.49** ile en başarılı performansa GR3 veri grubu ile ulaşılmıştır. Geliştirilen model CSE-WSE seçim fonksiyonu ile belirlenen 6 değişken kullanılarak **%96** tahmin doğruluğunda Osmaniye ili KGI'nı başarılı bir şekilde tahmin etmiştir. GR3 veri grubu için geliştirilen modelin ağ eğitim performans grafiği Şekil 5.30'da verilmiştir. Eğitim sürecinde geliştirilen en iyi modele 821 epok sonucunda **MSE = 0.0028** değeri ile ulaşmıştır.

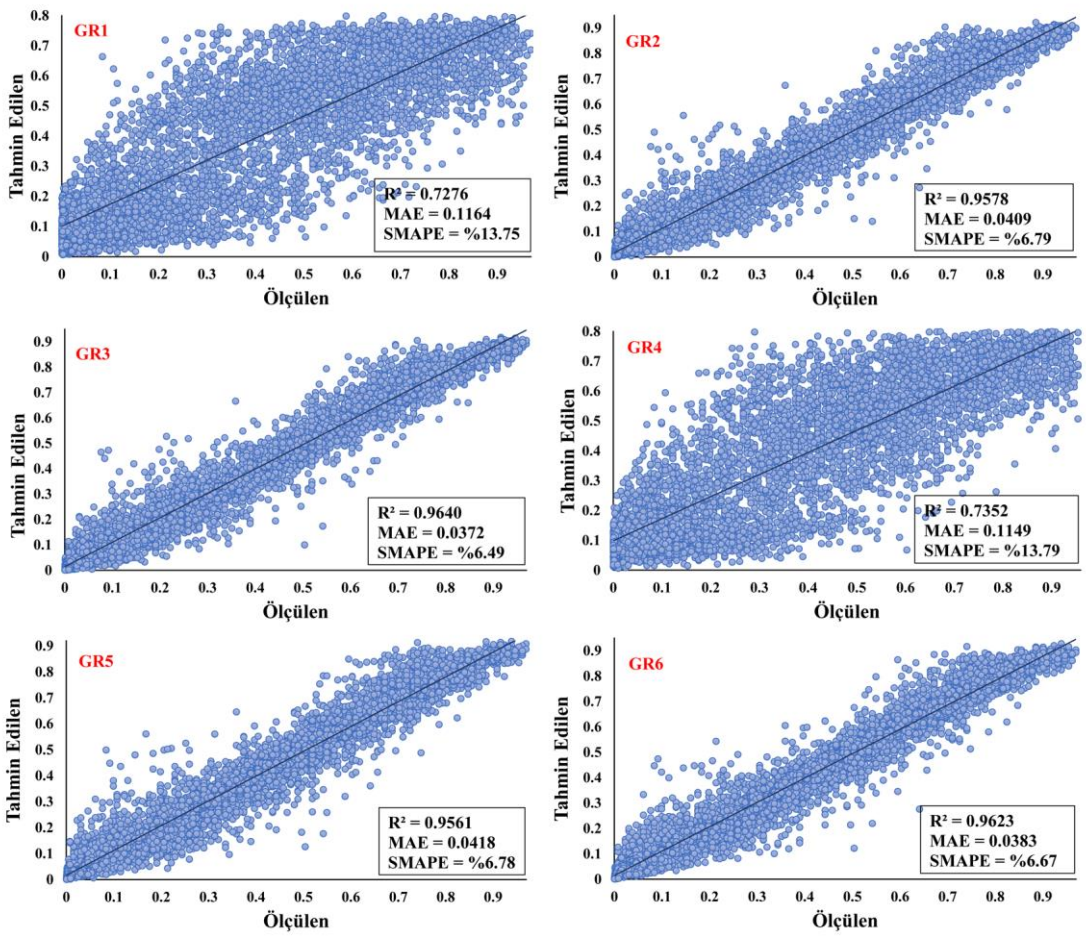
Çizelge 5.8. Osmaniye ili veri gruplarına göre geliştirilen en başarılı modellerin eğitim ve test tahmin sonuçları

		Modellerin Eğitim Performansı						**Modellerin Test Performansı**			
Veri Grubu	**Gizli Katman Nöron Sayısı**	**Eğitim Tekrar Sayısı**	**Eğitim Verisi MSE**	**Test Verisi MSE**	**Doğrulama Verisi MSE**	**Model MSE**	**Model R**	**RMSE**	**MAE**	**SMAPE (%)**	**R^2**
GR1	18	5	0.0213	0.0209	0.0211	0.0214	0.8580	0.1482	0.1164	13.75	0.7276
GR2	36	1	0.0030	0.0032	0.0029	0.0031	0.9806	0.0583	0.0409	6.79	0.9578
GR3	***45***	***5***	***0.0024***	***0.0030***	***0.0026***	***0.0028***	***0.9840***	***0.0539***	***0.0372***	***6.49***	***0.9640***
GR4	31	3	0.0207	0.0207	0.0210	0.0209	0.8607	0.1461	0.1149	13.79	0.7352
GR5	37	5	0.0033	0.0034	0.0033	0.0035	0.9792	0.0595	0.0418	6.78	0.9561
GR6	42	3	0.0027	0.0030	0.0028	0.0028	0.9826	0.0552	0.0383	6.67	0.9623

Hedef il için GR1-GR6 arasında geliştirilen tahmin modellerinin ölçülen ve tahmin edilen değerleri arasındaki dağılım grafikleri Şekil 5.31'de verilmiştir. Grafiklerden GR2, GR3, GR5 ve GR6 veri grupları tahmin modellerinin güneş ışınımını yüksek düzeyli bir korelasyonda tahmin ettiği ve GR3 veri grubunun kısmen daha başarılı bir performansa ulaştığı anlaşılmaktadır. GR1 ve GR4 veri gruplarının KGI'nı tahmin etmede diğer veri gruplarına kıyasla kötü sayılabilecek bir performans seviyesine ulaştıkları görülmektedir.



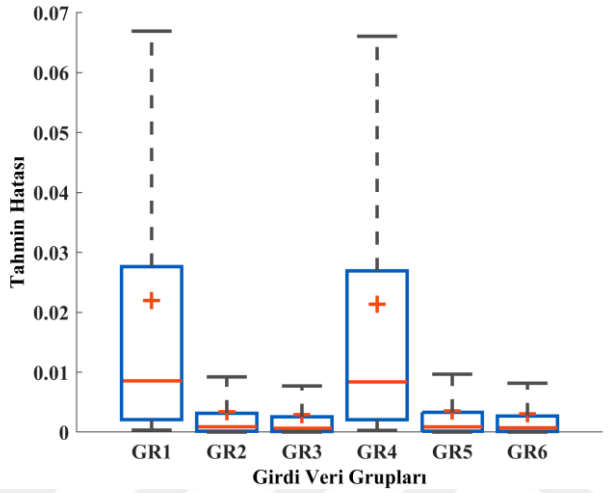
Şekil 5.30. Osmaniye ili için geliştirilen en iyi modelin ÇKİBSA performans grafiği



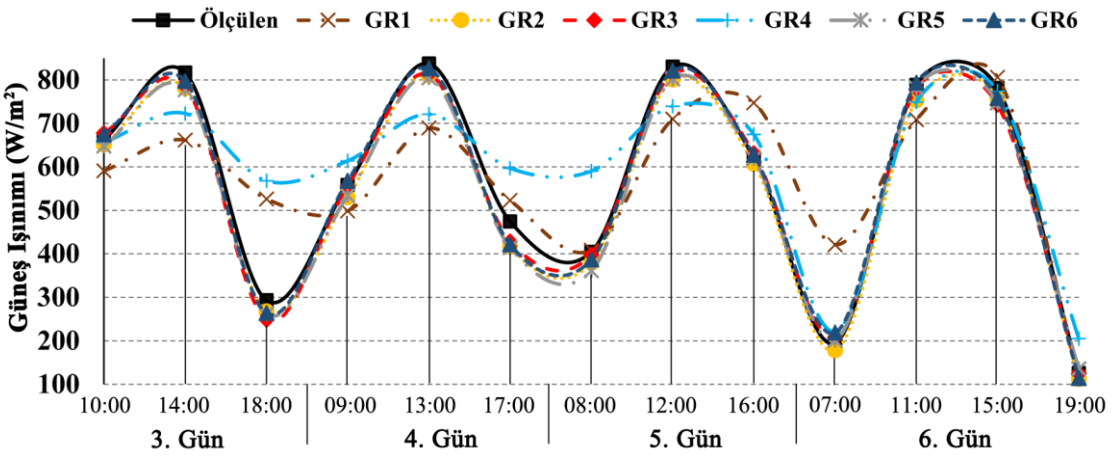
Şekil 5.31. Osmaniye ili test veri gruplarına göre geliştirilen ÇKİBSA modellerinin KGI tahmin dağılım grafikleri

Osmaniye ili GR1-GR6 veri gruplarına göre ölçülen ve tahmin edilen değerler arasındaki hata kutu grafiği Şekil 5.32'de verilmiştir. GR2, GR3, GR5 ve GR6 veri

grupları ile geliştirilen modellerinin hata oranlarının az olduğu, GR1 ve GR4 veri gruplarında ise hata oranlarının arttığı görülmektedir. Bunlar arasında GR3 ve GR6 veri grupları ile geliştirilen modelin en az hataya sahip olduğu anlaşılmaktadır. Tüm girdi veri grupları kullanılarak hesaplanan en başarılı modellerin ölçülen ve tahmin edilen değerleri arasındaki karşılaştırmalı performans grafiği Şekil 5.33'de verilmiştir.



Şekil 5.32. Osmaniye ili veri gruplarına göre ölçülen ve tahmin edilen değerler arasındaki hata kutu grafiği



Şekil 5.33. Osmaniye ili 2004 yılı Temmuz ayı dört günlük ölçülen ve tahmin edilen saatlik verilerin karşılaştırmalı performans grafiği

Karşılaştırmalı grafikte test girdi verilerinden rastgele seçilen 2004 yılının Temmuz ayının 3-6 günleri arası dört günlük saatlik verileri kullanılmıştır. GR2, GR3, GR5 ve GR6 veri grupları kullanılarak geliştirilen modellerin KGI'nı tahmin etmede oldukça iyi bir performans gösterdiği anlaşılmaktadır. VM süreçleri kullanılarak

oluşturulan özellik seçim gruplarının genel olarak hedef ilin güneş ışınımını modellemede yüksek bir tahmin başarısı sağladığı sonucuna varılmıştır.

5.1.9. ÇKİBSA tahmin modellerinin değerlendirilmesi

ÇKİBSA algoritması kullanılarak Akdeniz bölgesinin çalışma alanı olarak belirlenen 8 ilinin saatlik ortalama KGI potansiyellerini yüksek doğrulukta, amaca uygun ve yorumlanabilir bir şekilde tahmin edilmesi sağlanmıştır. Bölge çerisindeki her bir ilin KGI tahmin sonuçları ayrı ayrı hesaplanarak, Akdeniz bölgesinin sekiz ili tez çalışmasında kapsamlı olarak modellenmiştir. Her bir il için yapılan modelleme çalışmaları sonucunda güneş ışınımını en iyi tahmin eden sinir ağı modellerinin performans sonuçları Şekil 5.34'te verilmiştir.

Çizelge 5.5.9. Akdeniz bölgesi illerinin en başarılı KGI modellerinin tahmin performansı sonuçları

Çalışma Alanları	İstatistiksel Ölçekler				Veri Grubu
	RMSE	MAE	SMAPE (%)	R^2	
Adana	0.0544	0.0386	6.15	0.9542	GR3
Antalya	0.0516	0.0353	7.13	0.9546	GR3
Burdur	0.0514	0.0354	7.49	0.9650	GR3
Hatay	0.0579	0.0392	7.50	0.9565	GR7
Isparta	0.0536	0.0352	7.77	0.9488	GR3
K.Maraş	0.0508	0.0341	7.79	0.9656	GR2
Mersin	0.0481	0.0329	6.11	0.9698	GR3
Osmaniye	0.0539	0.0372	6.49	0.9640	GR3
Ortalama	0.0527	0.0360	7.05	0.9598	GR3

Çizelgeden her bir ilin en iyi modellerinin tahmin performansı sonuçlarının birbirlerinden farklı olduğu görülmektedir. Hesaplanan sonuçlarda her bir bölge için toplanan verilerin karakteristik özellikleri etkili olmuştur. Modelleme çalışmalarının hazırlık aşamasında VM süreçleri kullanılarak hedef iller için 6 farklı özellik seçim fonksiyonuna dayalı girdi veri grupları belirlenmiştir. Adana, Antalya, Burdur, Isparta Kahramanmaraş, Mersin, Osmaniye illeri için özellik seçim süreçleri sonunda GR1-GR5 arası, Hatay ili için GR1-GR6 arası veri grupları oluşturulmuştur. Süreç sonunda bazı özellik seçim fonksiyonlarında benzer sonuçlara ulaşıldığından bu gruplar birleştirilmiştir. GR1-GR5 arası özellik seçim gruplarının oluşturulduğu her bir il için GR6 ile tüm özelliklerin dahil olduğu bir değişken grubu eklenirken, GR1-GR6 arası

seçim gruplarına GR7 ile tüm girdi özellikleri dahil edilmiştir. Bu sayede özellik seçim süreçleri tabanlı geliştirilen tahmin modellerinin hem kendi aralarında hem de tüm girdi değişkenleri kullanılarak geliştirilen modeller arasındaki tahmin performansları değerlendirilmiştir.

Adana, Burdur, Hatay, Kahramanmaraş, Mersin ve Osmaniye illerinde GR2, GR3, GR5 ve GR6 veri grupları ile geliştirilen modeller genel olarak oldukça iyi performans gösterirken, Antalya ve Isparta illerinde GR2, GR3 ve GR6 veri gruplarının tahmin performansları daha başarılı bulunmuştur. Akdeniz bölgesinin tüm illerinde ClassAE (GR2), CSE (GR3), RAE (GR5) ve WSE (GR3, Hatay ilinde GR6) özellik seçim fonksiyonları kullanılarak oluşturulan veri grupları ile geliştirilen ÇKİBSA modellerinin KGI tahmin performansları oldukça iyi seviyede hesaplanırken, CfsSE (GR1) ve CorrAE (GR4) tabanlı veri gruplarının Akdeniz bölgesinin KGI'nı tahmin etmede arzulanan seviyede performans seviyesine ulaşamadığı anlaşılmıştır. GR6 ve GR7 veri grupları ile özellik seçim tabanlı girdi veri gruplarının tahmin performansları karşılaştırıldığında, Hatay ili hariç diğer tüm illerde özellik seçim gruplarının KGI'nın tahmin edilmesinde en iyi performans seviyesine ulaşıldığı gözlemlenmiştir. Adana'da 4 girdili, Antalya'da 5 girdili, Burdur'da 5 girdili, Isparta'da 7 girdili, Mersin'de 4 girdili ve Osmaniye'de 6 girdili GR3 girdi veri grubu ile geliştirilen modellerin tahmin performansları daha başarılı olurken, Kahramanmaraş ilinde 5 girdili GR2 veri grubu ile geliştirilen modelin güneş ışınım tahmin performans sonuçları daha iyi bulunmuştur. Akdeniz bölgesinde GR3 veri grubunun oluşturulmasında kullanılan CSE ve WSE özellik seçim fonksiyonları ile GR2 veri grubunun oluşturulmasında kullanılan ClassAE özellik seçim fonksiyonunun ÇKİBSA algoritması ile geliştirilen modellerin tahmin performansını üzerine oldukça etkili olduğu görülmüştür. CSE ve WSE özellik seçim fonksiyonlarının bölge üzerinde uygulama geliştirilen 8 farklı ilin altısında en etkili girdi değişkenlerinin belirlenmesinde oldukça etkili oldukları anlaşılmıştır.

VM süreçlerinin genel olarak ML tabanlı geliştirilen tahmin modellerinde uyumlu bir şekilde çalışarak tahmin kesinliğini arttırdığı gözlemlenmiştir. Bununla birlikte geliştirilen uygulamalarda veri havuzundaki anlamlı özelliklerin seçilmesi ile modelleme süresinin kısaldığı ve modellerin yorumlanabilmesini sağladığı sonucuna varılmıştır. Örneğin Adana ili için modelleme çalışmasında kullanılan 4 kategorik ve 5 meteorolojik girdi verisinden en iyi performans, Ay, Saat, S, GS verilerinin dahil olduğu GR3 veri grubu ile elde edilmiştir. Özellik sayısının azaltılması modelleme süresini kısaltmıştır. Aynı zamanda, Adana ili için tahmin modellerinde kullanılan

meteorolojik verilerden KGI üzerinde en etkili olanların S ve GS olduğu, diğer verilerin tahmin performansına önemli bir etkisinin olmadığı veya negatif yönde bir etkisinin olduğu çıkan sonuçtan açık bir şekilde yorumlanabilmektedir. 4 girdili GR3 veri grubunda en iyi performans SMAPE ölçeğine göre **%6.15** hesaplanırken, 9 girdili GR6 veri grubunda **%8.24** bulunmuştur. Benzer şekilde GR3 veri grubu ile R^2 ve MAE ölçeğine göre sırasıyla **0.9542** ve **0.0386** değerleri bulunurken, GR6 veri grubu ile **0.9527** ve **0.0402** değerleri hesaplanmıştır. GR6 veri grubu ile geliştirilen modelin tahmin performansı GR3 veri grubuna oldukça yakın bulunmuştur. Ancak çalışmada faydalanılan tüm özelliklerin kullanılması modelleme süresini oldukça uzatmış, aynı zamanda modelin sağlıklı bir şekilde yorumlanmasını engellemiştir. Girdi özelliklerinin tamamının kullanıldığı çalışmalarda veriden ziyade algoritma tahmin performansı hedef alındığından, girdi değişkenlerinin geliştirilen modelin performansında eşit seviyede etkili olduğu varsayılmaktadır (Long ve ark., 2014). Modelleme çalışmalarının yapısına aykırı olan bu yöntemin üstesinden gelebilmek adına VM süreçlerinden faydalanılarak hibrit bir yöntem benimsenmiştir.

Akdeniz bölgesi hedef alınarak yapılan çalışmada, aslında her bir ilin sınırlarına özgü bir modelleme çalışması yapılmıştır. Bu sebeple bir il için geliştirilen tahmin modelinin bir başka ilin girdi verilerine uyarlanması mümkün olamamaktadır. Akdeniz bölgesinde belirlenen 8 ilin tamamı hedef alınarak KGI'nn tahmin edilmesi için toplanan verilerin birleştirilerek eğitim ve test verilerinin yeniden düzenlenmesi ile modelleme çalışmalarında kullanılması amaçlanmıştır. Bu sayede bir çalışmada il bazında ve bölgesel olarak model performansları farklı açılardan değerlendirilmiş olacaktır.

Bölge genelini kapsayan bir çalışma yapmak için ilk olarak daha önce çalışmada faydalanılan meteorolojik ve kategorik 9 girdi verisine, her bir il için ayrı ayrı enlem, boylam ve yükseklik coğrafik verileri dahil edilerek 12 farklı girdi değişkeni belirlenmiştir. Ardından tüm veriler VM sürçlerine sokularak 6 farklı özellik seçim fonksiyonu temelli GR1-GR6 arasında girdi veri grupları oluşturulmuştur. GR7 ile tüm özelliklerin dahil olduğu veri grubu çalışmaya dahil edilmiştir. Seçim süreçleri sonunda belirlenen tüm özelliklerin dahil olduğu veri grupları Çizelge 5.10'da verilmiştir. Eğitim veri setinde 140,672 saatlik veri dahil edilirken, test veri setinde 46,864 saatlik veri ayarlanmıştır. ÇKİBSA algoritması kullanılarak her bir il için uygulanan modelleme sürecinden faydalanılarak geliştirilen modeller yardımıyla her bir veri grubu için en başarılı KGI tahmin sonuçları hesaplanarak Çizelge 5.11'de gösterilmiştir. Modellerin

eğitim performansları değerlendirildiğinde GR3, GR5, GR6 ve GR7 veri grubu kullanılarak geliştirilen modellerin tahmin performanslarının R istatistik ölçeğine göre yaklaşık **0.97** seviyelerinde ve birbirlerine oldukça yakın hesaplandığı görülmektedir. Test performansları değerlendirildiğinde en başarılı modelin **SMAPE = %7.78, MAE = 0.0341** ve $\mathbf{R^2}$ **= 0.9587** değerleri ile GR3 girdi veri grubu kullanılarak hesaplanmıştır. Ayrıca, GR2, GR5 ve GR6 özellik seçim veri grupları kullanılarak geliştirilen modellerin KGI tahmin performansları iyi düzeyde hesaplanırken, GR1 ve GR4 veri grupları KGI'nı başarılı bir şekilde tahmin etmede yetersiz kalmışlardır. GR7 veri grubu ile geliştirilen modelin tahmin performansı GR3 veri grubuna oldukça yakın bulunmuştur. Ancak, modelleme süreci yüksek performanslı bir bilgisayar kullanılmasına ve Matlab uygulama yazılımının çok çekirdekli paralel işleme desteğinden faydalanılmasına rağmen oldukça uzun bir süreyi aldığından kullanılması pek efektif görülmemektedir.

Çizelge 5.10. Akdeniz bölgesi için belirlenen özellik seçim grupları

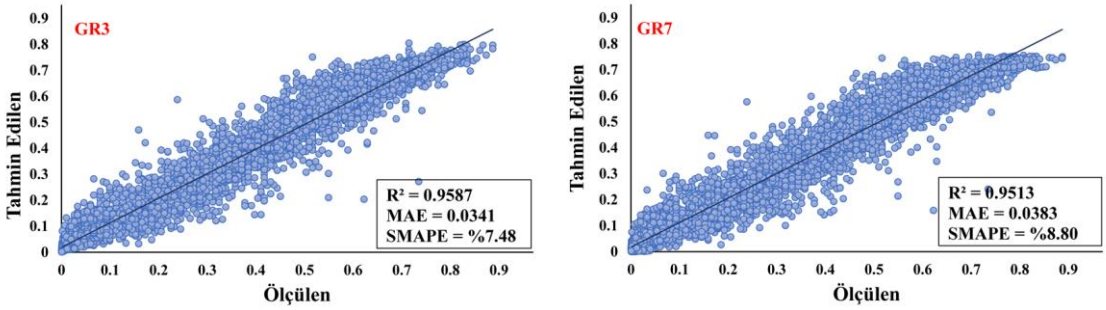
Girdi Veri Grupları	Özellik Seçim Fonksiyonları	Seçilen Özellikler	Toplam Özellik Sayısı
GR1	CfsSE	S, N, GS	3
GR2	ClassAE	Ay, Saat, S, N, RH, GS	6
GR3	CSE	Yıl, Ay, Saat, Enlem, Boylam, S, GS	7
GR4	CorrAE	Gün, Enlem, Yükseklik, S, RH, GS	6
GR5	RAE	Ay, Saat, Enlem, Boylam, Yükseklik, S, GS	7
GR6	WSE	Ay, Gün, Saat, Boylam, GS	5
GR7	-	Yıl, Ay, Gün, Saat, Enlem, Boylam, Yükseklik, S, B, N, RH, GS	12

Çizelge 5.11. Akdeniz bölgesi veri gruplarına göre geliştirilen en başarılı modellerin eğitim ve test tahmin sonuçları

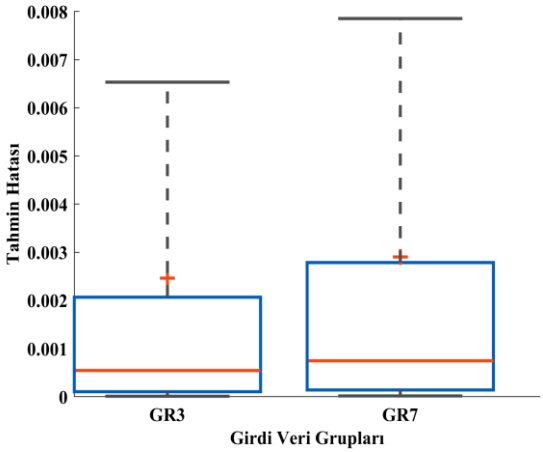
		Modellerin Eğitim Performansı						Modellerin Test Performansı			
Veri Grubu	Gizli Katman Nöron Sayısı	Eğitim Tekrar Sayısı	Eğitim Verisi MSE	Test Verisi MSE	Doğrulama Verisi MSE	Model MSE	Model R	RMSE	MAE	SMAPE (%)	R^2
GR1	31	3	0.0184	0.0185	0.0184	0.0187	0.8309	0.1352	0.1053	14.41	0.6936
GR2	49	5	0.0043	0.0043	0.0043	0.0044	0.9627	0.0658	0.0479	9.12	0.9274
GR3	***49***	***5***	***0.0024***	***0.0025***	***0.0025***	***0.0026***	***0.9791***	***0.0496***	***0.0341***	***7.48***	***0.9587***
GR4	50	1	0.0161	0.0162	0.0160	0.0161	0.8538	0.1266	0.0977	13.91	0.7313
GR5	41	5	0.0035	0.0034	0.0034	0.0035	0.9703	0.0588	0.0416	8.89	0.9422
GR6	42	2	0.0034	0.0035	0.0033	0.0035	0.9706	0.0586	0.0413	8.78	0.9425
GR7	49	5	0.0029	0.0029	0.0029	0.0028	0.9754	0.0539	0.0383	8.80	0.9513

Akdeniz bölgesinin 7 girdili GR3 ve 12 girdili GR7 veri grupları için geliştirilen en başarılı tahmin modellerinin ölçülen ve tahmin edilen değerleri arasındaki dağılım

grafikleri Şekil 5.34'de ve hata kutu grafikleri Şekil 5.35'te verilmiştir. Her iki grafikten bölgenin KGI'nı daha iyi tahmin eden modelin GR3 girdi veri grubu ile geliştirildiği açık bir biçimde anlaşılmaktadır.



Şekil 5.34. Akdeniz bölgesi GR3 ve GR7 veri gruplarına göre geliştirilen ÇKİBSA modellerinin KGI tahmin dağılım grafikleri



Şekil 5.35. Akdeniz bölgesi GR3 ve GR7 veri gruplarına göre ölçülen ve tahmin edilen değerler arasındaki hata kutu grafiği

Akdeniz bölgesinin genelini kapsayan modelleme sonuçları ile 8 ilin ayrı ayrı tahmin modellerinin değerlendirilmesi yapıldığında benzer bazı sonuçlara ulaşılmıştır. Çizelge 5.9'da R^2 ölçeğine göre en yüksek performans **0.9698** değeriyle Mersin ilinde, en düşük performans **0.9488** ile Isparta ilinde hesaplanmış, tüm illerin tahmin ortalaması **0.9598** bulunmuştur. Çizelge 5.11'de Akdeniz bölgesinin genelini kapsayan modelleme çalışmasında GR3 girdi veri grubuyla elde edilen en iyi tahmin performansı R^2 ölçeğine göre **0.9587** hesaplanmıştır. Benzer şekilde MAE ve SMAPE ölçeklerine göre 8 ilin tahmin ortalaması sırasıyla **0.0360** ve **%7.05** bulunurken, bölgenin genelini kapsayan çalışmada sırasıyla **0.0341** ve **%7.48** hesaplanmıştır. Bölge üzerinde yapılan

iki farklı çalışmada elde edilen sonuçların birbirlerine benzeştiği açık bir şekilde anlaşılmaktadır. Ayrıca VM süreçleri sonunda KGI üzerinde diğer özellik fonksiyonlarına kıyasla en etkili girdi özelliklerinin belirlenmesini sağlayan özellik seçim fonksiyonlarının her iki çalışmada da CSE ve WSE olduğu görülmektedir. Bununla birlikte VM süreci destekli ÇKİBSA ML algoritması kullanılarak geliştirilen tahmin modellerinin genel olarak oldukça iyi bir performans seviyesine ulaştığı sonucuna varılmıştır.

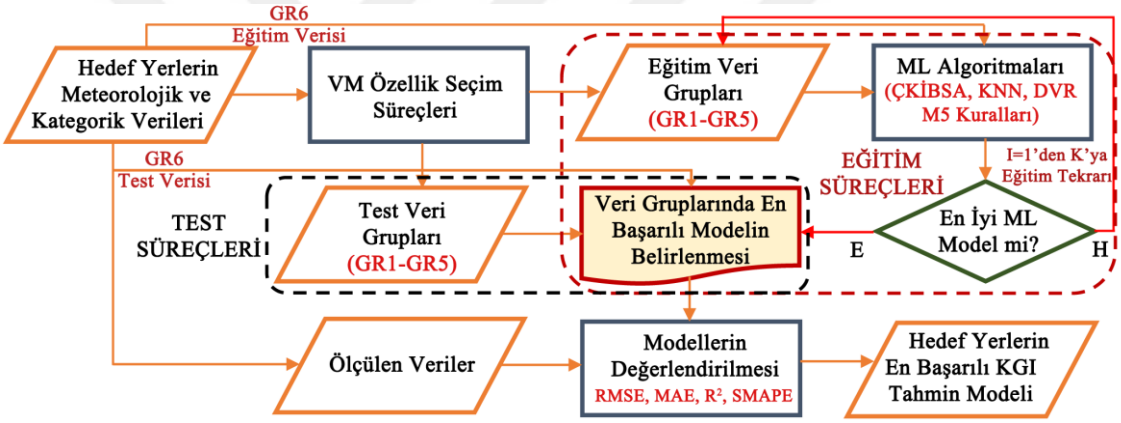
5.2. Farklı ML Algoritmaları ile Geliştirilen Modellerin Karşılaştırmalı Analizi

Tek bir ML algoritmasının tahmin gücünden yararlanılarak VM süreçleri sonunda düzenlenen girdi veri grupları yardımıyla Akdeniz bölgesi 8 ilinin ve bölge genelinin KGI tahmin performansları farklı açılardan değerlendirilmiştir. Veri-tahmin ilişkisinin hedef alındığı ve ÇKİBSA algoritmasının kullanıldığı bir önceki bölümdeki çalışmalarda, kapsamlı ve detaylı bir tahmin analizi gerçekleştirilmiştir. Seçilen algoritmanın modelleme kabiliyetine bağlı olarak tahminler yapıldığında, maksimum tahmin başarısı ile seçilen algoritmada kullanılacak veri grupları arasında kuvvetli bir ilişki bulunmaktadır. Bu bölümde veri-tahmin ilişkisinin değerlendirilmesine paralel olarak, algoritma-tahmin ilişkisinin değerlendirilmesi amaçlanmıştır. ÇKİBSA algoritmasının yanında, K-NN, M5 kuralları ve DVR ML algoritmaları kullanılarak Akdeniz bölgesinin aynı enleminde bulunan Kahramanmaraş ve Isparta illerinin 2002-2006 yılları arasında ölçülen meteorolojik ve kategorik verilerine göre saatlik ortalama KGI'nın karşılaştırmalı analizi amaçlanmıştır.

VM süreçleri sonunda özellik seçim fonksiyonlarının çıktı veri gruplarına göre her bir ML algoritması için tahmin modelleri geliştirilmiştir. Tahmin modelleri hem kendi aralarında hem de kullanılan algoritma temelinde değerlendirilerek karşılaştırılmıştır. Bununla birlikte, diğer çalışmalardan farklı olarak bu bölümde yapılan çalışmada, birbirlerine benzer bir yapıda olan klasik DVR algoritması ile DVM/DVR tabanlı LibSVM uygulaması kullanılarak tahmin modelleri geliştirilmiş ve performansları karşılaştırılarak en iyi ML yönteminin çalışmada kullanılması kararlaştırılmıştır.

Genellikle güneş ışınım tahmini üzerinde yapılan çalışmalarda, birçok bölge ve coğrafik yerin farklı veri türlerinden yararlanılmasına rağmen, aynı enlem üzerinde ve benzer coğrafik özellikler gösteren yerlerin verileri kullanılarak KGI tahmin

sonuçlarının değerlendirilmesi pek uygulanan bir yöntem değildir. Enlemin, güneşlenme süresi ve güneş ışınlarının geliş açısı üzerindeki etkisinden hareketle, (Darhmaoui ve Lahjouji, 2013), coğrafi bir bölgenin birbirinden oldukça uzak ve aynı enlem noktalarında yıllık güneş ışınım değerlerinin benzer bir seviyede olduğu ve optimum eğim açısı ile hedef enlem değeri arasında kuvvetli bir ilişkinin olduğunu hesaplamışlardır. (Ahlgren ve ark., 2018) ise, parabolik oluk toplayıcılarının bulunduğu enlem ve doğrudan normal ışınım değerleri arasında doğru orantılı ilişki bulunması sebebiyle yıllık verim ile enlem arasındaki ilişkiyi vurgulamışlardır. Bu sebeple, hedef bölge üzerinde aynı enlem koordinatlarına sahip yerler seçilmiş ve yüksek doğrulukta KGI tahmini için farklı türden ML algoritmaları kullanılarak karşılaştırmalı bir çalışma yapılmıştır. Kahramanmaraş ve Isparta illerinin farklı ML algoritmalarına göre saatlik ortalama KGI tahmin süreçlerinin akış şeması Şekil 5.36'da verilmiştir.

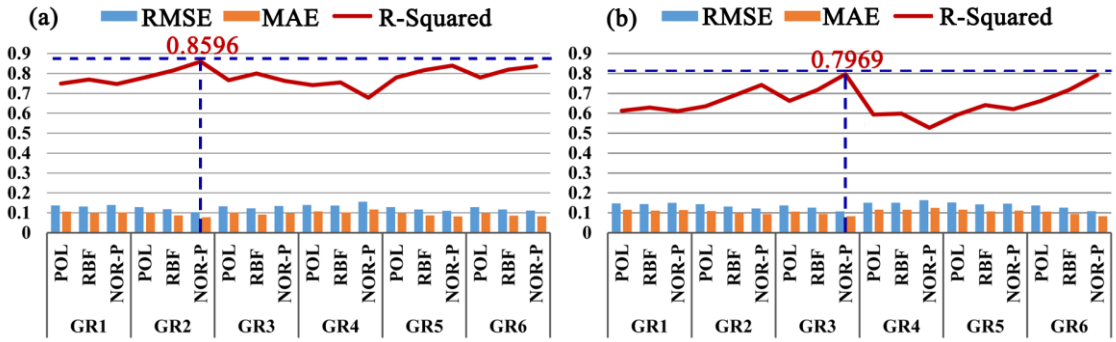


Şekil 5.36. Farklı ML algoritmaları kullanılarak KGI tahmin süreçlerinin şematik gösterimi

5.2.1. DVR algoritması ile geliştirilen modellerin tahmin bulguları

Güneş ışınım tahmin çalışmalarında YSA tabanlı ML algoritmalarından sonra en fazla yararlanılan ve performansı yüksek ML algoritması DVR'dur. Her iki ilin girdi veri grupları kullanılarak yapılan klasik DVR tabanlı tahmin modellerinde hesaplanan sonuçlar oldukça düşük bulunmuştur. Bu nedenle, klasik DVR tahmin sonuçlarının, DVR tabanlı bir başka yöntem olan LibSVM kütüphanesi ile değerlendirilmesine karar verilmiştir. LibSVM akademik literatürde bilinen bir yöntem olduğu ve KGI tahmininde nadir kullanıldığı için tercih edilmiştir. Her iki yöntemde de kullanıcı tarafından tanımlanan C (karmaşıklık ve maliyet parametresi), epsilon (hata parametresi) ve Nu (C

yerine kullanılan parametre) parametrelerinin seçiminde çok sayıda deneme uygulamaları yapılarak en uyumlu kombinasyonlar belirlenmiş ve veri grupları temelinde en başarılı tahmin modelleri geliştirilmiştir. Klasik SVR ile POL, NOR-P ve RBF çekirdek fonksiyonları kullanılarak her bir il için geliştirilen 18 farklı modelin performans sonuçları Şekil 5.37'de gösterilmiştir. SVR ile Kahramanmaraş ilinde R^2 ölçeğine göre **0.6786-0.8596** aralığında, Isparta ilinde **0.5273-0.7969** aralığında tahmin sonuçları elde edilmiştir.

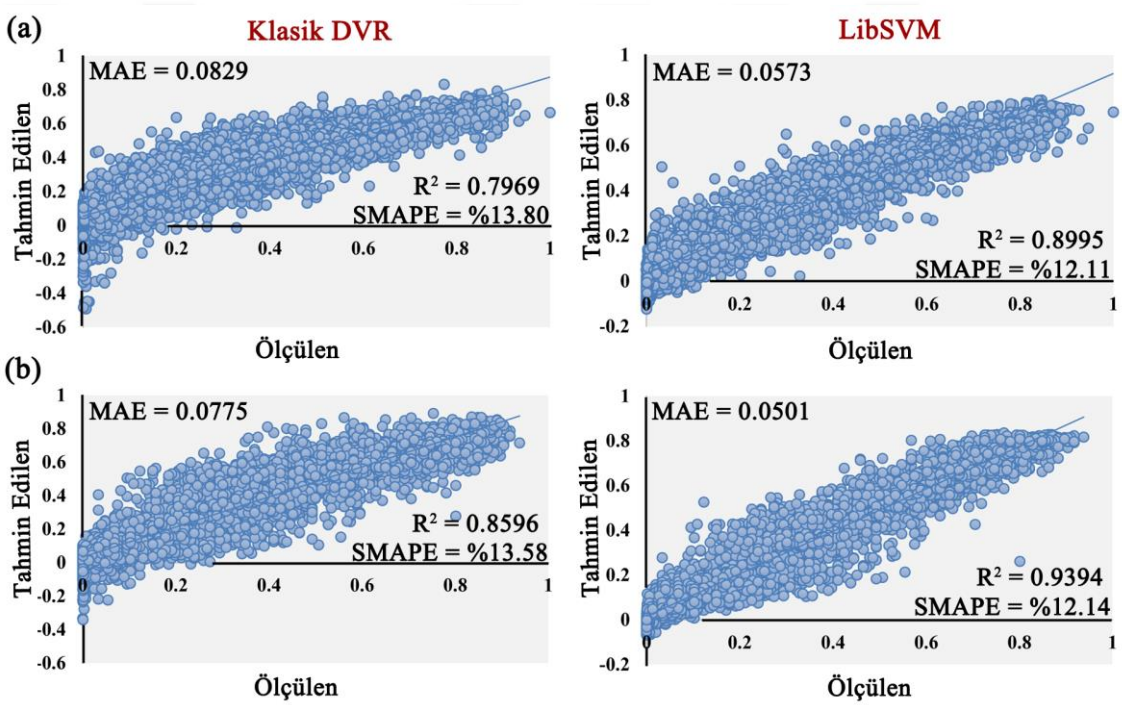


Şekil 5.37. Klasik DVR ile geliştirilen modellerin (a) Kahramanmaraş (b) Isparta ileri girdi veri gruplarına göre tahmin performansları.

LibSVM ile her bir ilin veri grupları için 12 farklı tahmin modeli geliştirilmiştir. Uygulamada iki farklı regresyon tabanlı DVR algoritmasında RBF çekirdek fonksiyonu kullanılarak geliştirilen KGI tahmin modellerinin istatistiksel sonuçları Çizelge 5.12'de verilmiştir. Nu-SVR ile geliştirilen modeller E-SVR'den daha başarılı olmuştur. Veri grupları temelinde bakıldığında her iki il için geliştirilen GR2, GR3, GR5 ve GR6 tahmin modellerinin sonuçları birbirlerine oldukça yakın ve başarılı bulunmuştur. GR2 veri grubu ile geliştirilen model Kahramanmaraş hedef bölesinde RMSE, MAE, SMAPE ve R^2 ölçeklerine göre sırasıyla **0.0675, 0.0501, %12.14** ve **0.9394** değerleri ile en başarılı tahmin performansına ulaşırken, benzer şekilde Isparta ilinde GR3 veri grubu kullanılarak geliştirilen modelde sırasıyla **0.0752, 0.0573, %12.11** ve **0.8995** değerleri hesaplanmıştır. Her iki ilde GR1 ve GR4 veri grupları ile geliştirilen modellerin tahmin sonuçları en düşük bulunmuştur. Şekil 5.38'de çalışmada kullanılan iki farklı DVR yöntemleri ile geliştirilen en başarılı modellerin belirlenen illere göre karşılaştırmalı dağılım grafikleri verilmiştir.

Çizelge 5.12. Her iki ilin girdi veri gruplarına göre LibSVM'ye göre geliştirilen en başarılı modellerin tahmin sonuçları

Hedef İller		Kahramanmaraş				Isparta			
Veri Grupları	DVM Tipi	RMSE	MAE	SMAPE (%)	R^2	RMSE	MAE	SMAPE (%)	R^2
GR1	E-SVR	0.1271	0.1003	15.11	0.7876	0.1431	0.1153	15.46	0.6372
	Nu-SVR	0.1257	0.0960	14.97	0.7893	0.1428	0.1144	15.44	0.6377
GR2	E-SVR	0.0742	0.0584	13.43	0.9278	0.0816	0.0637	12.64	0.8818
	Nu-SVR	**0.0675**	**0.0501**	**12.14**	**0.9394**	0.0765	0.0580	12.05	0.8961
GR3	E-SVR	0.0765	0.0604	13.69	0.9229	0.0815	0.0643	12.95	0.8821
	Nu-SVR	0.0697	0.0523	12.42	0.9352	**0.0752**	**0.0573**	**12.11**	**0.8995**
GR4	E-SVR	0.1300	0.1026	15.17	0.7781	0.1495	0.1193	15.66	0.6050
	Nu-SVR	0.1287	0.0984	14.97	0.7792	0.1491	0.1185	15.64	0.6057
GR5	E-SVR	0.0780	0.0616	13.77	0.9197	0.0849	0.0673	12.28	0.8728
	Nu-SVR	0.0717	0.0540	12.59	0.9315	0.0827	0.0621	11.77	0.8792
GR6	E-SVR	0.0802	0.0631	13.80	0.9146	0.0827	0.0645	12.52	0.8790
	Nu-SVR	0.0756	0.0576	12.96	0.9237	0.0779	0.0594	12.15	0.8925



Şekil 5.38. Klasik DVR ve LibSVM'ye göre geliştirilen en başarılı modellerin (a) Isparta, (b) Kahramanmaraş illerine göre karşılaştırmalı dağılım grafikleri

Dağılım grafiklerinden, her iki benzer yöntemden LibSVM ile geliştirilen modellerinin en iyi KGI tahmin sonuçları, klasik DVR'ye göre daha başarılı bulunmuştur. Bu sebeple ML algoritmalarının karşılaştırmalı değerlendirilmesinde LibSVM'nin tahmin sonuçlarının kullanılmasına karar verilmiştir. Her iki yöntemde geliştirilen modellerin DVR ML algoritmasına dayanmasına rağmen, aynı veri gruplarında farklı sonuçlar vermesi yazılımda kullanılan algoritma mimarisine ve benzerde olsa bazı parametrelerin farklı kullanılmasına bağlanmıştır.

5.2.2. K-NN algoritması ile geliştirilen modellerin tahmin bulguları

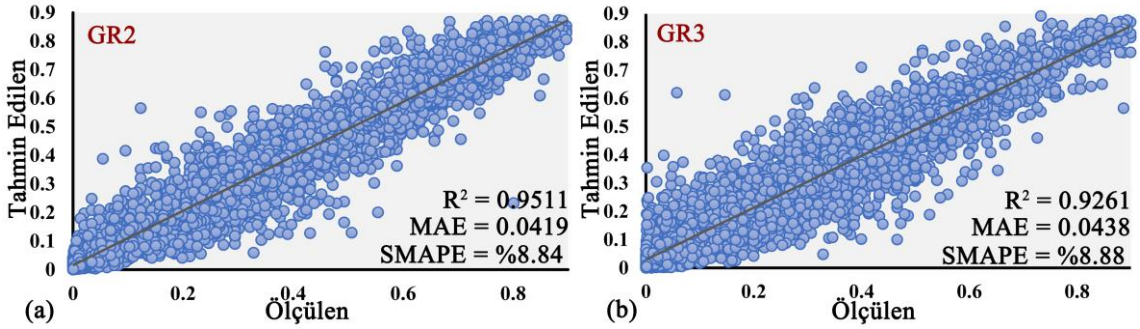
K-NN ML algoritması ile 1-10 arasında belirlenen altı farklı K komşuluk katsayısına göre her bir ilin seçilen girdi veri grupları kullanılarak 36 farklı tahmin modeli geliştirilmiştir. Kullanıcı tanımlı K parametresi ile her veri grubunda geliştirilen en başarılı iki modelin tahmin performans sonuçları Çizelge 5.13'te verilmiştir.

Çizelge 5.13. Hedef yerlerin veri gruplarına göre K-NN algoritması kullanılarak geliştirilen en başarılı iki modelinin tahmin sonuçları

Hedef İller		**Kahramanmaraş**				**Isparta**			
Veri Grubu	**Komşuluk Sayısı (K)**	**RMSE**	**MAE**	**SMAPE (%)**	**R^2**	**RMSE**	**MAE**	**SMAPE (%)**	**R^2**
GR1	6	0.1221	0.0903	14.25	0.8022	0.1447	0.1140	15.53	0.6274
	10	0.1205	0.0896	14.20	0.8069	0.1434	0.1131	15.42	0.6338
GR2	4	0.0618	0.0427	8.91	0.9490	0.0692	0.0481	9.34	0.9151
	6	**0.0605**	**0.0419**	**8.84**	**0.9511**	0.0693	0.0485	9.38	0.9150
GR3	4	0.0631	0.0437	9.08	0.9470	**0.0646**	**0.0438**	**8.88**	**0.9261**
	6	0.0631	0.0439	9.13	0.9469	0.0648	0.0450	9.17	0.9258
GR4	6	0.1288	0.0956	14.44	0.7798	0.1506	0.1157	15.42	0.5980
	10	0.1257	0.0937	14.30	0.7898	0.1463	0.1134	15.26	0.6194
GR5	4	0.0645	0.0446	9.37	0.9445	-	-	-	-
	6	0.0649	0.0448	9.32	0.9439	0.0820	0.0614	9.33	0.8805
	10	-	-	-	-	0.0811	0.0610	9.32	0.8829
GR6	3	0.0718	0.0487	10.34	0.9318	-	-	-	-
	4	0.0708	0.0482	10.32	0.9335	0.0728	0.0514	9.79	0.9065
	6	-	-	-	-	0.0730	0.0522	10.07	0.9062

K parametresinin tahmin modellerinde belirleyici bir özellik olduğu ancak her zaman artış yönünde bir doğru orantının bulunmadığı anlaşılmıştır. 10'un üzerindeki K parametresi ile geliştirilen tüm modellerde belirgin bir performans artışı gerçekleşmemiş ve modelleme süresi uzamıştır. İlgili tabloda, Kahramanmaraş ili için geliştirilen en başarılı model GR2 veri gurubu ile KGI'nı RMSE, MAE, R^2 ve SMAPE ölçeklerine göre sırasıyla **0.0605, 0.0419, 0.9511** ve **%8.84** ile tahmin etmiştir. Isparta ili için GR3 veri grubunda benzer şekilde sırasıyla **0.0646, 0.0438, 0.9261** ve **%8.88** değerleri hesaplanmıştır. Kahramanmaraş ilinde GR2, GR3 ve GR5 özellik seçim veri grupları ile geliştirilen modeller daha yüksek bir performans göstermişlerdir. GR6 tüm özelliklerin kullanıldığı veri grubu tahmin performansı özellik seçim gruplarının altında kalmıştır. Isparta ilinde ise GR2, GR3 ve GR6 veri grupları ile geliştirilen modellerin ölçülen gerçek değerlere %90'nın üzerinde bir tahmin doğruluğu ile benzeştiği hesaplanmıştır. Her iki ilde en kötü performans sonuçları GR1 ve GR4 veri gruplarında alınmış ve en başarılı tahmin sonuçlarının dağılım grafikleri Şekil 5.39'da verilmiştir.

SMAPE ölçeğine göre çalışmada faydalanılan illerin güneş ışınım tahminlerinin birbirlerine oldukça yakın olduğu ve benzeştiği görülmektedir.



Şekil 5.39. K-NN tabanlı en başarılı modellerin (a) Kahramanmaraş GR2, (b) Isparta GR3 girdi verilerine göre dağılım grafikleri

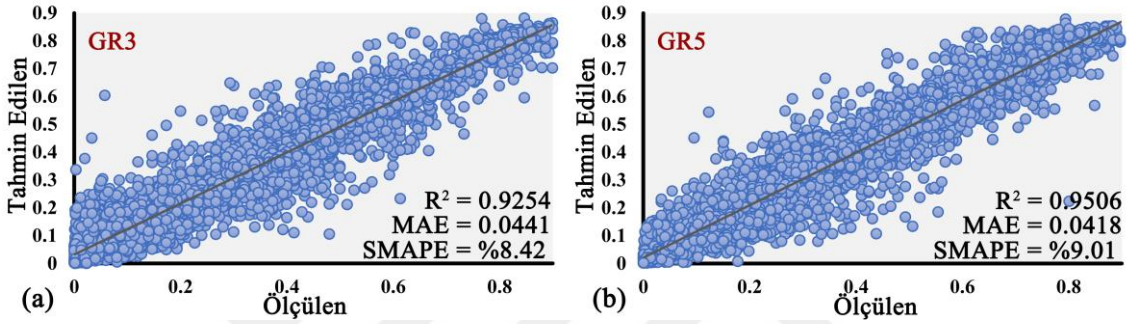
5.2.3. M5 kuralları algoritması ile geliştirilen modellerin tahmin bulguları

M5 kuralları algoritması ile hedeflenen illerin seçilen veri grupları kullanılarak, her bir il için altı farklı kural tabanlı tahmin modeli geliştirilmiştir. Geliştirilen modellerin tahmin performansları Çizelge 5.14'te verilmiştir. Kahramanmaraş ili için geliştirilen en iyi model, çalışmada faydalanılan diğer ML algoritmalarından farklı olarak GR5 özellik seçim veri grubu kullanılarak, Isparta ili ise benzer şekilde GR3 veri grubu kullanılarak tahmin edilmiştir. RMSE, MAE, R^2 ve SMAPE istatistiksel ölçeklerine göre Kahramanmaraş için en iyi modelin performansında sırasıyla **0.0610, 0.0418, 0.9506** ve **%9.01** değerleri elde edilmiştir. Isparta ili için benzer şekilde **0.650, 0.0441, 0.9254** ve **%8.42** değerleri bulunmuştur. Hedef illerde geliştirilen en başarılı modellerin dağılım grafikleri Şekil 5.40'ta verilmiştir. Grafiklere göre hedef illerin veri dağılımları ve performans ölçüm metriklerinin birbirlerine oldukça yakın bir seviyede hesaplandığı anlaşılmaktadır.

Çizelge 5.14. M5 Kuralları algoritması kullanılarak geliştirilen en başarılı modellerin tahmin sonuçları

Veri Grupları	Çalışma Alanı Olarak Belirlenen Hedef İller							
	Kahramanmaraş				Isparta			
	RMSE	MAE	SMAPE (%)	R^2	RMSE	MAE	SMAPE (%)	R^2
GR1	0.1216	0.0915	14.37	0.8026	0.1400	0.1114	15.46	0.6514
GR2	0.0625	0.0431	9.07	0.9479	0.0666	0.0447	8.77	0.9218
GR3	0.0625	0.0430	9.09	0.9479	**0.0650**	**0.0441**	**8.42**	**0.9254**
GR4	0.1254	0.0944	14.56	0.7903	0.1454	0.1140	15.44	0.6236
GR5	**0.0610**	**0.0418**	**9.01**	**0.9506**	0.0814	0.0593	9.37	0.8822
GR6	0.0614	0.0420	9.08	0.9500	0.0655	0.0445	8.43	0.9245

Kahramanmaraş ilinde, GR2, GR3, GR5 ve GR6 girdi veri gruplarının tahmin sonuçlarının gerçek ölçüm değerleri ile karşılaştırıldığında ortalama **%95** seviyelerinde birbirlerine oldukça yakın bir benzeşim sağlandığı görülmektedir. Isparta ilinde ise GR2, GR3 ve GR6 veri gruplarında ortalama **%92** seviyelerinde bir tahmin doğruluğu yakalanmıştır. Her iki ilin SMAPE, MAE ve RMSE ölçüm metrikleri de bu sonuçları doğrulamaktadır. Özellik seçim süreçleri sonunda KGI tahmin modellerinin tahmin doğruluğu artmış ve genel olarak veri setinde hedef çıktıyla uyumlu değişkenlerin belirlenmesini sağlamışlardır.



Şekil 5.40. M5 Kuralları tabanlı en başarılı modellerin (a) Isparta GR3, (b) Kahramanmaraş GR5 girdi gruplarına göre dağılım grafikleri

5.2.4. Hedef yerlerin tahmin bulgularının değerlendirilmesi

Birden fazla ML algoritmasından yararlanılmasının ilk ve temel şartı algoritma-tahmin performanslarının karşılaştırmalı değerlendirilmesini sağlamaktır. Bunun yanında, VM süreçleri sonunda belirlenen girdi veri gruplarının tamamının çalışmada kullanılması ile tez çalışmasının ana hedefi olan veri-tahmin sonuçlarının analizi farklı algoritmalar üzerinde gerçekleştirilmiştir. Normal şartlarda seçilen birden fazla algoritmanın tahmin performansları değerlendirilirken, çalışmada düzenlenen tek bir veri seti kullanılmakta ve tüm uygulamalar bu veri seti üzerinde modellenmektedir.

DVR tabanlı LibSVM, K-NN ve M5 Kuralları ML algoritmalarının tahmin sonuçları ile Isparta ve Kahramanmaraş illerinde daha önce ÇKİBSA algoritması ile hesaplanan KGI tahmin sonuçlarının beraber değerlendirmeye alınması kararlaştırılmıştır. Bu sayede çalışmada yararlanılan dört farklı ML algoritmasının performansları farklı yönlerden analiz edilebilecektir. Kullanılan tüm ML algoritmalarında, tahmin doğruluğunu arttırmak ve her veri grubunda en başarılı

modelleri seçmek amacıyla yapılan deneme çalışmalarının haricinde, ÇKİBSA, K-NN, M5 Kuralları ve DVR tabanlı LibSVM ile sırasıyla 3000, 72, 12 ve 24 farklı tahmin modelleri geliştirilmiştir. KGI tahmin performansları değerlendirildiğinde, Kahramanmaraş ilinde, özellik seçim süreçleri sonunda belirlenen GR2 (Ay, Saat, S, N, GS) ve GR5 (Ay, Saat, S, N, RH, GS) veri grupları ile geliştirilen modellerin tahmin doğruluğu daha iyi bulunmuştur. Isparta ilinde ise GR3 (Yıl, Ay, Gün, Saat, B, S, GS) veri grubu ile geliştirilen modeller daha yüksek performans seviyesine ulaşmışlardır. Her iki ilin güneş ışınım tahmininde kullanılan ML algoritmalarına göre en iyi performans gösteren modellerinin istatistiksel karşılaştırmaları Çizelge 5.15'de verilmiştir.

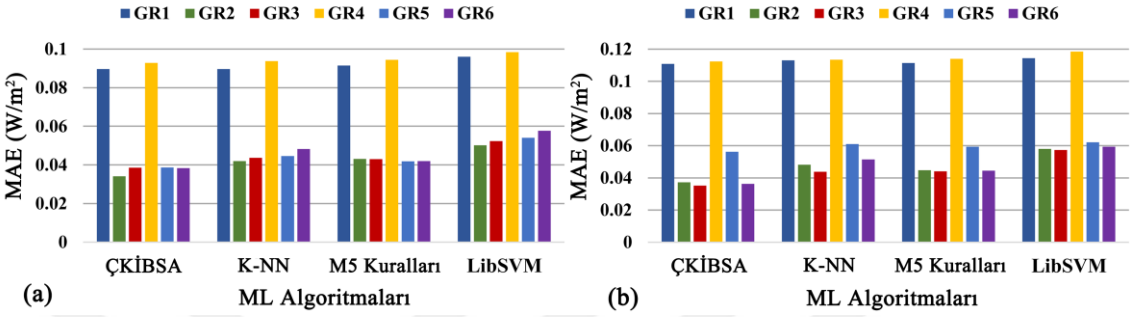
Çizelge 5.15. Her iki ilin farklı ML algoritmaları kullanılarak geliştirilen en başarılı KGI tahmin modellerinin karşılaştırması

İstatistiksel Ölçekler	Çalışmada Seçilen Hedef Yerler							
	Kahramanmaraş				Isparta			
	ÇKİBSA	K-NN	M5	LibSVM	ÇKİBSA	K-NN	M5	LibSVM
RMSE	**0.0508**	0.0605	0.0610	0.0675	**0.0536**	0.0646	0.0650	0.0752
MAE	**0.0341**	0.0419	0.0418	0.0501	**0.0352**	0.0438	0.0441	0.0573
R^2	**0.9656**	0.9511	0.9506	0.9394	**0.9488**	0.9261	0.9254	0.8995
SMAPE (%)	**7.79**	8.84	9.01	12.14	**7.77**	8.88	8.42	12.11
Veri Grupları	**GR2**	GR2	GR5	GR2	**GR3**	GR3	GR3	GR3

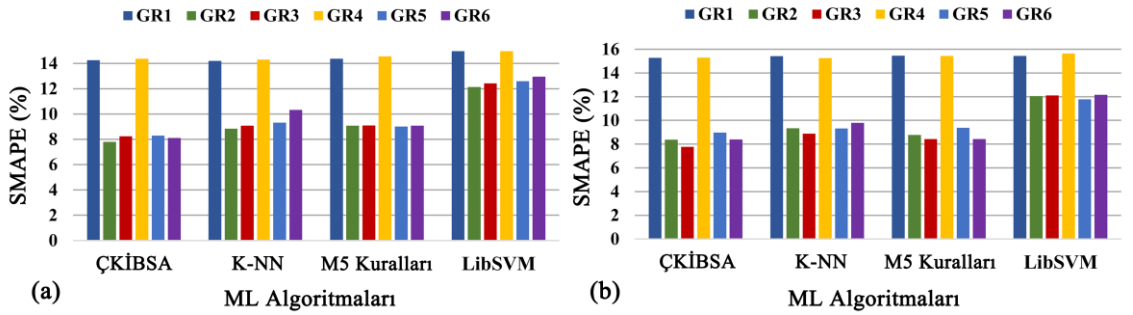
Çalışmada faydalanılan istatistiksel ölçeklere göre ÇKİBSA algoritması, her iki ilde güneş ışınımını diğer algoritmalara göre daha başarılı bir şekilde tahmin etmiştir. Bununla birlikte, her bir il için K-NN ve M5 Kuralları algoritması ile benzer tahmin sonuçlarına ulaşılırken, LibSVM geliştirilen tahmin modellerinde en düşük performans değerleri elde edilmiştir. ÇKİBSA algoritması ile Kahramanmaraş ve Isparta illerinde SMAPE'ye göre sırasıyla **%7.79** ve **%7.77**, K-NN algoritması ile sırasıyla **%8.84** ve **%8.88** ve LibSVM algoritması ile sırasıyla **%12.14** ve **%12.11** KGI tahmin sonuçlarına ulaşılmıştır. SMAPE'ye göre, çalışmada seçilen her iki ilin güneş ışınım tahmin sonuçlarının birbirlerine çok yakın bir seviyede olması enlem ve birtakım coğrafik özelliklerinin benzerliklerine bağlanmıştır.

KGI tahmin çalışmalarında, eldeki tüm girdi verilerin kullanıldığı GR6 veri grubu ile geliştirilen tahmin modellerin nihai performansı, özellik seçim süreci sonunda belirlenen GR2, GR3 ve GR5 gibi veri grupları ile geliştirilen modellerin nihai sonuçlarından daha düşük çıkmaktadır. Yapılan çalışmalarda, hedef iller için kullanılan tüm veri gruplarının 4 ML algoritmasına göre MAE ve SMAPE performans grafikleri

Şekil 5.41 ve 5.42'de verilmiştir. VM süreçlerinde faydalanılan özellik seçim fonksiyonlarının geliştirilen tahmin modellerinin performansına pozitif bir katkısının olduğu açık bir şekilde görülmektedir. VM destekli ML algoritmalarının KGI'nı tahmin etmede modelleme çalışmalarını hızlandırdığı, tahmin doğruluklarını arttırdığı ve modellerin yorumlanmasını sağladıkları sonucuna varılmıştır.



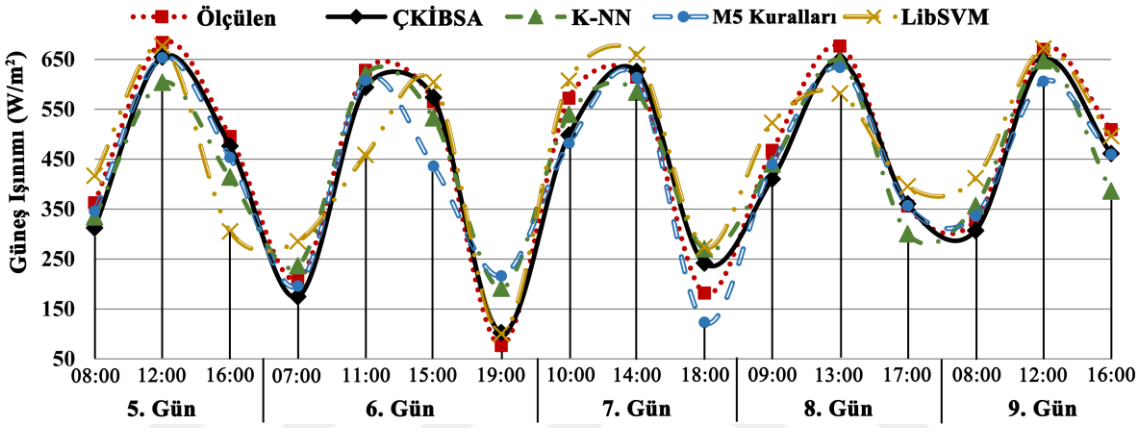
Şekil 5.41. Veri gruplarına göre ML algoritmaları kullanılarak geliştirilen en iyi modellerin (a) Kahramanmaraş, (b) Isparta illeri için MAE performans karşılaştırmaları



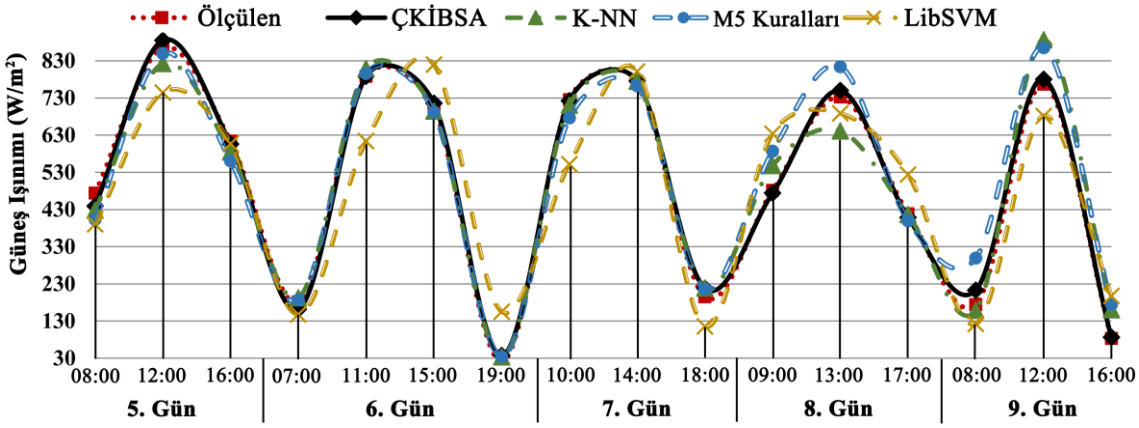
Şekil 5.42. Veri gruplarına göre ML algoritmaları kullanılarak geliştirilen en iyi modellerin (a) Kahramanmaraş, (b) Isparta illeri için SMAPE performans karşılaştırmaları

Kahramanmaraş ve Isparta illeri için geliştirilen KGI tahmin modelleri, genel olarak hedef yerlerin güneş enerjisi potansiyelini iyi bir şekilde tahmin etmiştir. Ancak, GR1 ve GR4 veri grubuyla yapılan çalışmalar her iki ilde tahmin performansı en düşük olan girdi veri gruplarını temsil eder. Bunun sonucunda, meteorolojik ve kategorik girdi veri setine uygulanan özellik seçim fonksiyonlarından CfsSE ve CorrAE'nın en iyi girdi verilerini belirlemede yetersiz kaldığı sonucuna varılmıştır. En başarılı özellik seçim fonksiyonları Kahramanmaraş için ClassAE ve WSE, Isparta için CSE ve WSE'dir. Isparta ilinin CSE ve WSE özellik seçim fonksiyonları ile belirlenen 7 girdi verisi kullanılarak 4 ML algoritması ile geliştirilen en iyi modellerinin güneş ışınım tahminleri

ile gerçek ölçüm sonucu karşılaştırmaları Şekil 5.43'te, benzer şekilde Kahramanmaraş ili için ClassAE özellik seçim fonksiyonu için 5 girdili geliştirilen en iyi modellerin karşılaştırmaları Şekil 5.44'te verilmiştir. Yapılan karşılaştırmalarda test verisinden rastgele seçilen 2004 yılı Temmuz ayının 5-9 günleri arası beş günlük saatlik girdi verileri kullanılmıştır. Her gün için rastgele seçilerek oluşturulan test verilerinin saat dilimlerinde, ölçülen ve tahmin sonuçlarının karşılaştırmasında, Kahramanmaraş ili için geliştirilen saatlik ortalama KGI tahmin modellerinin Isparta iline göre daha iyi güneş ışınımını tahmin ettiği görülmektedir.



Şekil 5.43. Isparta ili 2004 yılı Temmuz ayı beş günlük girdi verileri kullanılarak ölçülen ve tahmin edilen 4 farklı ML modellerinin karşılaştırmalı performans grafiği



Şekil 5.44. Kahramanmaraş ili 2004 yılı Temmuz ayı beş günlük girdi verileri kullanılarak ölçülen ve tahmin edilen 4 farklı ML modellerinin karşılaştırmalı performans grafiği

Literatürde, saatlik ortalama KGI'nın ML algoritmaları kullanılarak tahmin edilmesi amacıyla yapılan önceki çalışmalar ile Isparta ve Kahramanmaraş illerinin 4 ML algoritması kullanılarak tahmin edilen en başarılı modeli arasındaki karşılaştırma

Çizelge 5.16'da sunulmaktadır. Çizelgeden görüldüğü üzere, geliştirilen en başarılı modeller, yaygın olarak, bu çalışmada olduğu gibi sinir ağı temellidir. Önerilen VM tabanlı ML tahmin modellerinin performansı, yapılan önceki çalışmalardan daha iyi oldukları anlaşılmaktadır.

Çizelge 5.16. Literatürde saatlik ortalama KGI tahmin çalışmaları ile mevcut çalışmanın karşılaştırılması

ML Algoritması	En İyi Algoritma	Zaman Aralığı	Çalışma Alanı	Yazar	Ölçek		
					RMSE	MAE	R^2
YSA	YSA	2001-2007	La Serena (Şili)	(Lazzús ve ark., 2011)	-	-	0.9437
YSA	YSA	02 Şubat-31 Mayıs	Cezayir	(Hasni ve ark., 2012)	0.172	2.9971	0.9999
MLP	MLP	2009-2012	Fez (Fas)	(Ihya ve ark., 2014)	-	-	0.8896
MLP-NARX	NARX	2010-2014	Fes (Fas)	(Loutfi ve ark., 2017a)	-	-	0.95
AdaBoost, LR, KNN, CART, DVR, ANN, RD Regresyon	RD Regresyon	2013-2015	Güney Kore	(Kim ve ark., 2019)	577.5	-	0.705
YSA	YSA	2006-2010	Ajaccio (Korsika)	(Notton ve ark., 2012)	12.43	19.17	0.998
MARS-ANN-LR	YSA	2010-2016	Hong Kong	(Li ve ark., 2019)	0.270	0.194	0.918
ÇKİBSA-KNN-M5-LibSVM	**ÇKİBSA**	**2002-2006**	**K.Maraş-Isparta**	**Mevcut Çalışma**	**0.0341**	**0.0508**	**0.9656**

6. SONUÇLAR VE ÖNERİLER

6.1 Sonuçlar

Enerji, sosyo-ekonomik hayatın sürdürülebilmesi için en önemli yapı taşıdır. Günümüzde ülkeler için güvenli, yeterli miktarda, ucuz ve temiz enerji üretmek en öncelikli sorundur. Yaygın olarak kullanılan petrol temelli enerji kaynaklarının tükenebilir oluşu, çevreye olan olumsuz etkisi, dışa bağımlılığın artması ve ülkeler arasındaki politik sorunlar sebebiyle alternatif enerji kaynaklarına yönelim artmaktadır. Bu sebeple, hâlihazırda kullanılan enerji kaynaklarına ilaveten daha farklı enerji kaynaklarına yöneliminin gerekliliği büyük önem taşımaktadır.

Yenilenebilir enerji kaynakları arasında yer alan güneş enerjisi son yıllarda birçok ülke tarafından en fazla yatırımın yapıldığı önemli ve sürekli bir enerji türüdür. Gittikçe artan önemi sayesinde güneş enerjisi çevre dostu teknolojiler ile temiz enerji vadetmektedir. Ancak, bir bölgeye güneş tabanlı bir sistem kurulmadan önce mutlaka ilgili bölgenin güneş enerji potansiyelinin hesaplanması gereklidir. Meteorolojik olarak ölçümü en uygun yöntem olsa da cihazların hassas ve pahalı olmasından dolayı maliyetlidir ve uzmanlık gerektirmektedir. Ayrıca ölçümler için uzun zaman dilimlerine ihtiyaç duyulmaktadır. Bu sebeple, dünya genelinde en yaygın ve kullanışlı hesaplama yöntemi, çeşitli meteorolojik ve coğrafik veriler kullanılarak ilgili yerin güneş ışınım potansiyelini tahmin etmektir. Güneş ışınımın tahmin edilmesinde deneysel, istatistiksel ve yapay zekâ tabanlı birçok yöntem ve teknik kullanılmaktadır.

Ülkemiz dünya ile mukayese edildiğinde güneş enerjisi bakımından önemli bir potansiyele sahiptir. Bölgeler bazında bakıldığında Akdeniz Bölgesi güneş enerjisi potansiyeli olarak Doğu Anadolu Bölgesi'nden sonra ikinci sıradadır. Bu yüzden yapılması planlanan çalışmanın Akdeniz Bölgesi temelli olması kararlaştırılmıştır. Bölge içerisinde belirlenen Adana, Antalya, Burdur, Isparta, Hatay, Kahramanmaraş, Mersin ve Osmaniye illeri çalışma alanı olarak belirlenmiştir.

YZ tabanlı uygulamalar, son zamanlarda güneş ışınımının tahmin edilmesinde en fazla yararlanılan yöntemlerin başında gelmektedirler. Bilgisayar bilimi içerisinde oldukça geniş bir uygulama alanı olan YZ, birçok bilimsel ve teknolojik çalışmada kullanılmaktadır. Bu yöntemler içerisinde en yaygın ve güncel olanı ML algoritmalarıdır. İçerisinde denetimli, yarı denetimli, denetimsiz ve takviyeli öğrenme tekniğine dayanan birçok algoritma barındıran ML yöntemleri genel olarak veri türü ve

yapısından fazlasıyla etkilenen hassas uygulamalar olarak bilinmektedirler. Birçok çalışmada göz ardı edilen veri-tahmin ilişkisi nedeniyle hedeflenen ölçüde tahmin performanslarına ulaşmak mümkün olamamaktadır. Bir tahmin sonucuna ulaşılır ancak daha iyi performansa ulaşılabilirliği kafada bir soru işareti olarak kalmaktadır. Güneş ışınımın tahmini üzerine yapılan literatür taramalarında, bu ilişki kısmen göz önüne alınmıştır. Genellikle ML algoritmasının modelleme gücü ön plana alınarak veri üzerinde temel ve basit hazırlıklar ile modellemeler gerçekleştirilmiştir. Yapılan bazı çalışmalarda ise seçilen tek bir algoritmanın modelleme gücünden ziyade, birden fazla algoritmanın aynı veri seti üzerindeki tahmin performanslarının karşılaştırılması hedeflenmiştir. Bu tip çalışmalarda ana amaç veri-tahmin ilişkisinden ziyade algoritma-tahmin ilişkisini değerlendirmektir. Bu nedenle veri üzerinde yapılacak hazırlıklar büyük oranda çok fazla önemsenmemektedir.

Literatür taramalarında tek bir ML algoritmasına dayalı KGI tahmin çalışmalarında eldeki tüm girdi değişkenlerinin kullanıldığı gözlemlenmiştir. Tahmin modellerinde algoritmanın mimari yapısı ve parametreleri üzerinde değişiklikler yapılarak farklı tahmin modelleri geliştirilmiştir. Ancak veriye dayalı model çeşitliliğini arttırarak en başarılı modellerin seçimi son zamanlarda daha fazla tercih edilmektedir. Bu amaçla veri setinden değişkenlerin seçiminde farklı yöntem ve tekniklerin izlendiği görülmüştür. Bazı çalışmalarda deneme yanılma yoluyla girdi değişkenleri seçilirken, bazılarında temel yöntem ve tekniklerle en iyi özellikler belirlenmektedir. Veri seti içinden amaca uygun en iyi değişkenlerin seçilmesi, her değişkenin aynı oranda çıktı üzerinde etkili olamayacağı fikrine dayanmaktadır. Bazı özellikler çıktı üzerinde pozitif etki yaratırken, bazılarının hiçbir etkisi olmamakta veya modeli negatif olarak etkilemektedir. Ayrıca eldeki tüm girdi özelliklerini kullanmak modelleme süresini uzattığı gibi modelin yorumlana bilirliğini de engellemektedir. Bununla birlikte girdi özelliklerinin tamamının kullanılması parametrik olmayan makine öğrenme algoritmalarının aşırı uyum göstermesine yol açabilmektedir. Güneş ışınımına da pozitif ve negatif anlamda etki eden veya edebilecek birçok veri türü bulunduğundan en optimal verilerin seçimi bu açıdan önemli bulunmuştur.

Bu tez çalışmasında, Akdeniz bölgesinin çalışma alanı olarak belirlenen yerlerinin, KGI'nı etkili ve verimli bir şekilde tahmin edebilmesi için veri-tahmin ve algoritma-tahmin ilişkisine dayanan bir yöntem izlenmiştir. Literatürde var olan ML tabanlı KGI tahmin çalışmaları bir çalışmada birleştirilerek her iki yöntemin uygulanması sağlanmıştır. Çalışma alanı olarak seçilen yerlerde VM süreçlerine

dayanan ML tabanlı tahmin modelleri geliştirilerek hem veri-tahmin hem algoritma-tahmin ilişkisi değerlendirilmiştir. Bu sayede VM süreçlerinin veri ön işlem prosesleri güneş tabanlı bir çalışmada ayrıntılı bir biçimde uygulanmış ve birden fazla ML algoritması kullanılarak geliştirilen modelleme süreçlerindeki KGI'nın tahmin kesinliğine etki seviyeleri analiz edilmiştir. Ayrıca herhangi bir yer için hesaplanan tahmin performansı tek başına bir ölçüt olmayıp, modelleme süresi ve modelin yorumlanması gibi farklı etmenlerde kıyaslanmıştır.

Bu amaçla ilk olarak Akdeniz bölgesinde belirlenen 8 il için MGM'den 2002-2006 yılları arasında ölçülen saatlik ortalama meteorolojik ve coğrafik veriler talep edilmiştir. Toplamda 6 meteorolojik ve 3 coğrafik veri türü alınmıştır. Bu veri setine 4 adet kategorik veri türü dahil edilmiştir. Her bir il için tahmin modelleri ayrı ayrı geliştirileceğinden coğrafik veriler çalışmaya dahil edilmemiş, sadece bölge geneli geliştirilen modelleme çalışmasında veri setine dahil edilmiştir. VM ön işlem süreçlerini kapsayan temizleme, birleştirme ve dönüştürme adımlarını kullanarak ham veriler üzerinde detaylı işlemler yapıp il bazında her bir girdi özelliği için 23,442 adet saatlik düzenlenmiş veri elde edilmiştir. Uzun bir zaman alan bu süreçte Matlab ve Excel programlarında kodlar yazılarak bilgisayar ortamında veri düzenleme işlemlerinin otomatik olarak gerçekleştirilmesi sağlanmıştır. Ardından amaca uygun optimal değişkenlerin belirlenmesi için veri madenciliğinin indirgeme adımında faydalanılan özellik seçim süreçlerinden yararlanılmıştır. Bu amaçla açık kaynak kodlu WEKA programından yararlanılarak, filtre ve sarmalayıcı yaklaşımlara dayanan 6 özellik seçim fonksiyonu kullanılarak geliştirilen 29 farklı uygulama sonucunda güneş ışınımı üzerinde etkisi en fazla olan girdi değişkenleri belirlenmiş ve gruplandırılmıştır. Hatay ili hariç 7 ilin veri gruplarında CSE ve WSE özellik seçim fonksiyonlarında aynı değişkenler bulunduğundan, bu illerde özellik seçimine dayalı GR1-GR5 arası nihai veri girdi grupları oluşturulmuştur. Bu gruplara GR6 ile tüm değişkenlerin dahil olduğu bir veri grubu dahil edilmiştir. Hatay ilinde ise tüm değişkenlerin dahil olduğu grup GR7 olarak belirlenmiştir.

Veri-tahmin ilişkisinin değerlendirilmesi tez çalışmasının ana konusu olduğundan, ÇKİBSA ML algoritması kullanılarak her bir il için kapsamlı bir modelleme çalışması gerçekleştirilmiştir. Akdeniz bölgesinin çalışma alanı olarak seçilen 8 ili için düzenlenip belirlenen nihai veri grupları kullanılmıştır. Matlab R2017b yazılımıyla geliştirilen bir program kodu ile 1-50 arası gizli katman nöronu için, her sinir nöronunda 5 tekrarlı bir yapı tasarlanarak her veri grubu için eğitim ve test tahmin

sonuçlarının bilgisayar ortamında otomatik hesaplanması sağlanmıştır. Intel I5 2.5 Ghz işlemcili ve 8 GB bellek kapasiteli bir taşınabilir bilgisayarın performansı düşük kaldığından, paralel işleme fonksiyonu ile modelleme sürecini kısaltmak için performansı yüksek bir iş istasyonundan faydalanılmıştır. Her bir veri grubu için 250, her bir il için 1,500 ile 1,750 arasında tahmin modeli geliştirip hem birbirleri hem veri grupları arasında değerlendirilerek en iyi tahmin modelleri bulunmuştur. Akdeniz bölgesinde belirlenen 8 ilin 7'sinde en iyi performans gösteren KGI tahmin modelleri GR2 ve GR3 özellik seçim grupları ile hesaplanırken, birinde tüm girdi değişkenlerinin dahil olduğu veri grubu ile tahmin edilmiştir. ClassAE, CSE, RAE ve WSE özellik seçim fonksiyonları kullanılarak seçilen veri grupları bölge genelinde oldukça iyi performans seviyesine ulaşırken, CfsSE ve CorrAE'nın performansları tüm illerin tahmin modellerinde beklentinin altında kalmıştır. GR2, GR3, GR5 ve GR6 özellik seçim grupları ile geliştirilen modellerin tahmin performansının tüm illerde genel olarak oldukça iyi bir tahmin kesinliğine ulaştığı belirtilmiş ve en iyi tahmin sonuçlarına, Hatay ili hariç tüm illerde özellik seçim veri grupları ile ulaşılmıştır. Akdeniz bölgesinin genelini kapsayan bir çalışma da yapılmış tüm illerin hesaplama sonuçlarının ortalamalarına yakın sonuçlara ulaşılmıştır. Bu çalışmada da en iyi performans GR3 özellik seçim veri grubunda hesaplanmış ve 4 özellik seçim grubunda **%93**'ün üstünde bir tahmin doğruluğuna ulaşılmıştır. Tahmin performansları farklı istatistiksel ölçeklere değerlendirildiğinde bölge geneli yapılan çalışmanın, illerin ayrı ayrı hesaplanan ortalama tahmin sonuçları ile benzeştiği görülmüştür. Modelleme çalışmalarının tamamında eldeki tüm girdi değişkenlerini kullanmanın Hatay ili hariç tahmin performansına negatif bir etkisinin olduğu gösterilmiştir. VM süreçleri sonunda oluşturulan girdi veri gruplarının genel olarak tüm illerin KGI'nın tahmin edilesinde oldukça iyi bir performans seviyesine ulaştığı görülmüştür.

Yapılan bir başka çalışmada, ÇKİBSA, SVR tabanlı LibSVM, K-NN ve M5 Kuralları ML öğrenme algoritmaları kullanılarak aynı enlemde ve benzer coğrafik özellikler gösteren yerlerin algoritma-tahmin performansının değerlendirilmesi amaçlanmıştır. Bu amaçla Isparta ve Kahramanmaraş illerinin girdi veri grupları kullanılarak aynı zamanda VM süreçlerinin farklı algoritmalar üzerindeki etki seviyeleri değerlendirilmiştir. Modelleme çalışmaları sonunda her iki bölgede en iyi performans ÇKİBSA algoritması ile birbirlerine oldukça yakın hesaplanmıştır. En düşük performans DVR algoritması ile elde edilmiştir. K-NN ve M5 Kuralları algoritmalarının çıkan sonuçları her iki ilde birbirlerine oldukça yakın hesaplanmıştır. Özellik seçim

fonksiyonları, her iki ilde yapılan modelleme çalışmalarında, farklı ML algoritmaları üzerinde, en başarılı tahmin sonuçlarının belirlenmesinde etkili olmuştur. Genel olarak GR2, GR3 ve GR5 veri grupları kullanılarak geliştirilen modeller ML algoritmalarının tamamında tahmin performansı en iyi modeller olarak bulunmuştur. Ayrıca tahmin sonuçları karşılaştırıldığında, aynı enlemde birbirlerinden oldukça uzak ve benzer coğrafik özelliklere sahip yerlerin, güneş ışınım tahmin potansiyellerinin birbirlerine benzer olduğu söylenebilmektedir. Isparta ilinde 4 farklı ML algoritması ile geliştirilen modellerin tamamında GR3 veri grubunda en iyi tahmin sonuçları elde edilirken, Kahramanmaraş ilinde ÇKİBSA, LibSVM ve K-NN algoritmalarında GR2 ve M5 Kuralları algoritmasında GR5 özellik seçim veri grubu kullanılarak en iyi KGI tahmin sonuçlarına ulaşılmıştır.

Güneş ışınım tahmininde VM süreçlerinin ML algoritmaları ile beraber modelleme çalışmalarında kullanılması tahmin kesinliğinin yanında modelleme süresini kısalttığı ve modelin yorumlanabilmesini kolaylaştırdığı gözlemlenmiştir. Antalya, Burdur ve Mersin illerinde KGI üzerinde en etkili meteorolojik değişkenin **GS** olduğu, benzer şekilde Adana ve Osmaniye illerinde **S** ve **GS** olduğu görülürken, bazı değişkenlerin herhangi bir etkisi olmadığı ya da tahmin performansını negatif yönde etkilediği sonucuna varılmıştır. Isparta ilinde **B**, **S** ve **GS** değişkenleri en etkili meteorolojik veriler olarak belirlenirken, Kahramanmaraş ilinde ise **N**, **S** ve **GS** değişkenleri en etkili veriler olarak bulunmuştur. Akdeniz bölgesinin geneli üzerinde yapılan, 8 ilin tüm girdi verilerinin kullanıldığı modelleme çalışmasında KGI üzerinde en etkili özelliklerin **S** ve **GS** meteorolojik verileri olduğu anlaşılmıştır.

6.2 Öneriler

VM süreçleri kullanılarak MGM'den talep edilen verilerin işlenmesi sonunda oluşturulan veri grupları ile KGI'nın yüksek doğrulukta tahmini ve veri-tahmin ilişkisinin değerlendirilmesinde YSA mimarisine dayanan ÇKİBSA ML algoritmasından faydalanılmıştır. Kullanılan bu algoritmanın yerine farklı bir YSA mimarisine dayanan bir algoritma tercih edilebilir. Bu tez çalışmasında algoritma-tahmin performanslarının değerlendirilmesinde 4 farklı ML algoritmasından faydalanılmıştır. ML algoritmalarının sayısı arttırılarak tahmin modellerinin geliştirilmesi ile karşılaştırmalı tahminler yapılabilir. Ayrıca ML algoritmaları ile derin öğrenme algoritmaları bir arada kullanılarak performans sonuçları karşılaştırılabilir.

ML algoritmaları büyük ölçüde kullanılacak verinin yapısından etkilendiği için veriler üzerinde tahminler yapılmadan önce kapsamlı bir hazırlık aşamasından geçirilmesinin gerekliliği bu tez çalışmasında ayrıntılı bir biçimde ele alınmıştır. VM sürçlerinin KGI'nın tahmin performansını üzerindeki olumlu etkisinden hareketle, çalışmada kullanılan özellik seçim fonksiyonlarının dışında farklı fonksiyonlar ve yöntemlerle güneş ışınımı üzerinde en etkili girdi özellikleri belirlenerek tahmin performansları değerlendirilebilir. Ayrıca sarmalayıcı özellik seçim fonksiyonlarında sınıflayıcı olarak kullanılan 10 farklı ML algoritmasının dışında daha farklı algoritmalardan yararlanılabilir.

ML öğrenme tabanlı tahmin modellerinin geliştirilmesinde Matlab R2017b programından faydalanılmıştır. Açık kaynak kodlu Scikit-learn, TensorFlow, Keras gibi python programlama dili tabanlı kütüphaneler kullanılarak tahmin modelleri geliştirilip mevcut çalışma ile karşılaştırılabilir. Ayrıca, tahmin sonuçlarının değerlendirilmesinde 6 farklı istatistiksel metrikten yararlanılmıştır. Kullanılan istatistiksel ölçeklerin sayısı arttırılarak geliştirilen modellerin tahmin performansları farklı yönlerden değerlendirilebilir.

Son yıllarda ülkemizde güneş panelleri ve çevre ürünlerinin yerli imkanlarla üretiminin artması ve devletin bu yerli ürünlerin kullanımını teşvik etmesi ile güneş enerjisi yatırımları oldukça cazip hale gelmiştir. Bu sebeple, Akdeniz bölgesi hedef alınarak yapılan bu çalışma ülke genelini kapsayacak şekilde genişletilebilir. VM süreçleri farklı bölgelerin güneş tabanlı verilerini işleyerek KGI tahmin performansları mevcut çalışmaya göre değerlendirilebilir. Bu sayede Türkiye'nin güneş ışınım potansiyeli net bir şekilde ortaya çıkartılarak yapılan çalışmanın kapsamlı bir bilgisayar programı haline getirilmesi planlandığından bu alanda yatırım ve bilimsel çalışma yapacak kişi, kurum ve kuruluşların hizmetine sunularak var olan önemli bir eksikliğin giderilmesi sağlanabilir.

Bölge üzerinde herhangi bir yere kurulacak güneş enerjisi santralinin KGI tahmin sonuçlarına göre modellenmesinin yapılması ve fizibilitesinin hesaplanmasını sağlanabilir. Bu sayede yapılacak santralin ilk maliyetinin ve genel karlılık durumunun raporlanması sağlanarak yatırımcıya, bu konuda yatırım yapmayı düşünenlere ve akademik çevrelere çok büyük bir katkı sunabilir.

KAYNAKLAR

Abedin, Z., Barua, M., Paul, S., Akther, S., Chowdhury, R. ve Chowdhury, M. S. U., 2017, A model for prediction of monthly solar radiation of different meterological locations of Bangladesh using aritficial neural network data mining tool, *2017 International Conference on Electrical, Computer and Communication Engineering (ECCE)*, 692-697.

Ahlgren, B., Tian, Z., Perers, B., Dragsted, J., Johansson, E., Lundberg, K., Mossegård, J., Byström, J. ve Olsson, O., 2018, A simplified model for linear correlation between annual yield and DNI for parabolic trough collectors, *Energy Conversion and Management*, 174, 295-308.

Ahmad, A., Anderson, T. N. ve Lie, T. T., 2015, Hourly global solar irradiation forecasting for New Zealand, *Solar Energy*, 122, 1398-1408.

Akıl, Y., 2019, Güneş Işınım Tahmini için Farklı Tahmin Yöntemlerinin Karşılaştırılması: Güneydoğu Anadolu Bölgesi Örneği, Yüksek Lisans Tezi, *Batman Üniversitesi*, Batman, 74.

Al-Alawi, S. M. ve Al-Hinai, H. A., 1998, An ANN-based approach for predicting global radiation in locations with no direct measurement instrumentation, *Renewable Energy*, 14 (1), 199-204.

Al-Rousan, N., Al-Najjar, H. ve Alomari, O., 2021, Assessment of predicting hourly global solar radiation in Jordan based on Rules, Trees, Meta, Lazy and Function prediction methods, *Sustainable Energy Technologies and Assessments*.

Al Shamisi, M., Assi, A. ve Hejase, H., 2011, Using MATLAB to Develop Artificial Neural Network Models for Predicting Global Solar Radiation in Al Ain City - UAE, In, Eds, p.

Alam, S., Kaushik, S. C. ve Garg, S. N., 2006, Computation of beam solar radiation at normal incidence using artificial neural network, *Renewable Energy*, 31 (10), 1483-1491.

Alkan, A., 2014, Gezgin Satıcı Problemi İçin Veri Madenciliği Tabanlı Bir Model Önerisi, Doktora Tezi, *Kocaeli Üniversitesi*, Kocaeli, 100.

Almaraashi, M., 2018, Investigating the impact of feature selection on the prediction of solar radiation in different locations in Saudi Arabia, *Applied Soft Computing*, 66, 250-263.

Alsafadi, M., 2020, Hourly Global Solar Radıatıon Estımatıon Usıng Empırıcal And Machıne Learnıng Models In Eskişehir, Yüksek Lisans Tezi, *Anadolu Üniversitesi*, Eskişehir, 71.

Alsina, E. F., Bortolini, M., Gamberi, M. ve Regattieri, A., 2016, Artificial neural network optimisation for monthly average daily global solar radiation prediction, *Energy Conversion and Management*, 120, 320-329.

Angstrom, A., 1924, Solar and terrestrial radiation. Report to the international commission for solar research on actinometric investigations of solar and atmospheric radiation, *Quarterly Journal of the Royal Meteorological Society*, 50 (210), 121-126.

Asuero, A. G., Sayago, A. ve González, A. G., 2006, The Correlation Coefficient: An Overview, *Critical Reviews in Analytical Chemistry*, 36 (1), 41-59.

Awad, M. ve Khanna, R., 2015, Machine Learning, In: Efficient Learning Machines: Theories, Concepts, and Applications for Engineers and System Designers, Eds: Awad, M. ve Khanna, R., *Berkeley, CA*: Apress, p. 1-18.

Ayodele, T. R., Ogunjuyigbe, A. S. O. ve Monyei, C. G., 2016, On the global solar radiation prediction methods, *Journal of Renewable and Sustainable Energy*, 8 (2).

Azimi, R., Ghayekhloo, M. ve Ghofrani, M., 2016, A hybrid method based on a new clustering technique and multilayer perceptron neural networks for hourly solar radiation forecasting, *Energy Conversion and Management*, 118, 331-344.

Bayrak, M. ve Esen, Ö., 2014, TÜRKİYE'NİN ENERJİ AÇIĞI SORUNU VE ÇÖZÜMÜNE YÖNELİK ARAYIŞLAR, *Ataturk University Journal of Economics & Administrative Sciences*, 28 (3), 139-158.

Behrang, M. A., Assareh, E., Ghanbarzadeh, A. ve Noghrehabadi, A. R., 2010, The potential of different artificial neural network (ANN) techniques in daily global solar radiation modeling based on meteorological data, *Solar Energy*, 84 (8), 1468-1480.

Belaid, S. ve Mellit, A., 2016, Prediction of daily and mean monthly global solar radiation using support vector machine in an arid climate, *Energy Conversion and Management*, 118, 105-118.

Benghanem, M., Mellit, A. ve Alamri, S. N., 2009, ANN-based modelling and estimation of daily global solar radiation data: A case study, *Energy Conversion and Management*, 50 (7), 1644-1655.

Benmouiza, K. ve Cheknane, A., 2013, Forecasting hourly global solar radiation using hybrid k-means and nonlinear autoregressive neural network models, *Energy Conversion and Management*, 75, 561-569.

Besharat, F., Dehghan, A. A. ve Faghih, A. R., 2013, Empirical models for estimating global solar radiation: A review and case study, *Renewable and Sustainable Energy Reviews*, 21, 798-821.

Bou-Rabee, M., Sulaiman, S. A., Saleh, M. S. ve Marafi, S., 2017, Using artificial neural networks to estimate solar radiation in Kuwait, *Renewable and Sustainable Energy Reviews*, 72, 434-438.

Box, G. E. P. ve Jenkins, G. M., 1970, Time series analysis, Forecasting and Control, *San Francisco*, Holden-Day, p.

Casti, J. L., 2004, Synthetic thought, *Nature*, 427 (6976), 680-680.

Cengiz, M. ve Mamis, M., 2015, Thermal Solar Energy Use and Turkey Analysis, *Bitlis Eren University Journal of Science and Technology*, 5.

Chang, C.-C. ve Lin, C.-J., 2011, LIBSVM: A library for support vector machines, *ACM Transactions on Intelligent Systems and Technology*, 2 (3), Article 27.

Chen, J.-L., Liu, H.-B., Wu, W. ve Xie, D.-T., 2011, Estimation of monthly solar radiation from measured temperatures using support vector machines – A case study, *Renewable Energy*, 36 (1), 413-420.

Chiteka, K. ve Enweremadu, C. C., 2016, Prediction of global horizontal solar irradiance in Zimbabwe using artificial neural networks, *Journal of Cleaner Production*, 135, 701-711.

Cihan, P., 2018, Veri Madenciliği Yöntemleriyle Hayvan Hastalıklarında Teşhis, Prognoz ve Risk Faktörlerinin Belirlenmesi, Doktora Tezi, *Yıldız Teknik Üniversitesi*, İstanbul, 86.

Coulson, K., 2012, Solar and Terrestrial Radiation: Methods and Measurements, Elsevier Science, p.

Çelik, Ö., Teke, A. ve Yıldırım, H. B., 2016, The optimized artificial neural network model with Levenberg–Marquardt algorithm for global solar radiation estimation in Eastern Mediterranean Region of Turkey, *Journal of Cleaner Production*, 116, 1-12.

Dahmani, K., Dizene, R., Notton, G., Paoli, C., Voyant, C. ve Nivet, M. L., 2014, Estimation of 5-min time-step data of tilted solar global irradiation using ANN (Artificial Neural Network) model, *Energy*, 70, 374-381.

Darhmaoui, H. ve Lahjouji, D., 2013, Latitude Based Model for Tilt Angle Optimization for Solar Collectors in the Mediterranean Region, *Energy Procedia*, 42, 426-435.

de Freitas Viscondi, G. ve Alves-Souza, S. N., 2019, A Systematic Literature Review on big data for solar photovoltaic electricity generation forecasting, *Sustainable Energy Technologies and Assessments*, 31, 54-63.

del Campo-Ávila, J., Piliougine, M., Morales-Bueno, R. ve Mora-López, L., 2019, A data mining system for predicting solar global spectral irradiance. Performance assessment in the spectral response ranges of thin-film photovoltaic modules, *Renewable Energy*, 133, 828-839.

Doğaner, M. S., 2020, Akdeniz Bölgesi Coğrafyası, İstanbul Üniversitesi Akademik Veri Yönetim Sistemi, https://avesis.istanbul.edu.tr/sudogan/dokumanlar: [02.01.2021].

Dorvlo, A. S. S., Jervase, J. A. ve Al-Lawati, A., 2002, Solar radiation estimation using artificial neural networks, *Applied Energy*, 71 (4), 307-319.

EİGM, 2016, Türkiye Güneş Enerjisi Potansiyeli Atlası (GEPA) Türkiye, https://gepa.enerji.gov.tr/MyCalculator/: [07.08.2020].

El Mghouchi, Y., Chham, E., Zemmouri, E. M. ve El Bouardi, A., 2019, Assessment of different combinations of meteorological parameters for predicting daily global solar radiation using artificial neural networks, *Building and Environment*, 149, 607-622.

Erdoğan, E., 2006, Zaman Serilerinde ARIMA Modelleri, Yüksek Lisans Tezi, *Muğla Üniversitesi Fen Bilimleri Enstitüsü İstatistik ve Bilgisayar Bilimleri Anabilim Dalı*, 87.

Fadare, D. A., 2009, Modelling of solar energy potential in Nigeria using an artificial neural network model, *Applied Energy*, 86 (9), 1410-1422.

Famili, A., Shen, W.-M., Weber, R. ve Simoudis, E., 1997, Data preprocessing and intelligent data analysis, *Intelligent Data Analysis*, 1 (1), 3-23.

Fayyad, U. M., Piatetsky-Shapiro, G. ve Smyth, P., 1996, Advances in knowledge discovery and data mining, In, Eds: Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P. ve Uthurusamy, R., *Menlo Park, CA, USA*: American Association for Artificial Intelligence, p. 1-34.

Feng, Y., Gong, D., Zhang, Q., Jiang, S., Zhao, L. ve Cui, N., 2019, Evaluation of temperature-based machine learning and empirical models for predicting daily global solar radiation, *Energy Conversion and Management*, 198.

Ghimire, S., Deo, R. C., Downs, N. J. ve Raj, N., 2019, Global solar radiation prediction by ANN integrated with European Centre for medium range weather forecast fields in solar rich cities of Queensland Australia, *Journal of Cleaner Production*, 216, 288-310.

Ghofrani, M., Niromand, N., Azimi, R. ve Ghayekhloo, M., 2017, A novel data mining method for high accuracy solar radiation forecasting, *2017 North American Power Symposium (NAPS)*, 1-6.

Gibert, K., Sànchez-Marrè, M. ve Izquierdo, J., 2016, A survey on pre-processing techniques: Relevant issues in the context of environmental data mining, *AI Communications*, 29, 627-663.

Gielen, D., Boshell, F., Saygin, D., Bazilian, M. D., Wagner, N. ve Gorini, R., 2019, The role of renewable energy in the global energy transformation, *Energy Strategy Reviews*, 24, 38-50.

Guher, A. B., Tasdemir, S. ve Yaniktepe, B., 2020, Effective Estimation of Hourly Global Solar Radiation Using Machine Learning Algorithms, *International Journal of Photoenergy*, 2020, 1-26.

Gülen, Ö. Ç., 2014, Veri Madenciliği Teknikleri ile Üstün Yetenekli Öğrencilerin İlgi Alanlarının Analizi, Doktora Tezi, *Gazi Üniveristesi*, Ankara, 160.

Gülsoy, N., 2019, Kredi Skorlama Süreçlerinde Veri Madenciliği ve Bankacılık Sektöründe Bir Uygulama, Doktora Tezi, *Erciyes Üniversitesi*, Kayseri, 2019.

Gültürk, E., 2016, Veri Madenciliğinde Kullanılan Kestirim Yöntemlerinin Performanslarının Karşılaştırlması, Doktora Tezi, *Cumhuriyet Üniversitesi*, Sivas, 143.

Hagan, M. T., Demuth, H. B., Beale, M. H. ve Jesús, O. D., 2002, Neural network design 2nd Edition, *Orlando University*, p. 1012.

Hamdan, M. A., Abdelhafez, E. ve Ghnaimat, O., 2017, Prediction of Hourly Solar Radiation in Amman-Jordan by Using Artificial Neural Networks.

Hasni, A., Sehli, A., Draoui, B., Bassou, A. ve Amieur, B., 2012, Estimating Global Solar Radiation Using Artificial Neural Network and Climate Data in the South-western Region of Algeria, *Energy Procedia*, 18, 531-537.

Hassan, M. Z., Ali, M. E. K., Ali, A. B. M. S. ve Kumar, J., 2017, Forecasting Day-Ahead Solar Radiation Using Machine Learning Approach, *2017 4th Asia-Pacific World Congress on Computer Science and Engineering (APWC on CSE)*, 252-258.

Hastie, T., Tibshirani, R. ve Friedman, J., 2009, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, *NY*, Springer-Verlag, p.

Haykin, S. S., 1999, Neural Networks: A Comprehensive Foundation, Prentice Hall, p.

Hejase, H. A. N., Al-Shamisi, M. H. ve Assi, A. H., 2014, Modeling of global horizontal irradiance in the United Arab Emirates with artificial neural networks, *Energy*, 77, 542-552.

Huan, L. ve Lei, Y., 2005, Toward integrating feature selection algorithms for classification and clustering, *IEEE Transactions on Knowledge and Data Engineering*, 17 (4), 491-502.

IEC, 2018, Artificial intelligence across industries. Wang, Y. Switzerland, IEC White Paper AI**:** 98.

Ihya, B., Mechaqrane, A., Tadili, R. ve Bargach, M. N., 2014, Prediction of hourly and daily diffuse solar fraction in the city of Fez (Morocco), *Theoretical and Applied Climatology*, 120 (3-4), 737-749.

İnan, O., 2015, Veri Madenciliği Uygulamaları için Veri İndirgeme Algoritmalarının Geliştirilmesi ve Resim Madenciliğine Uygulanması, Doktora Tezi, *Selçuk Üniversitesi*, Konya, 164.

Inman, R. H., Pedro, H. T. C. ve Coimbra, C. F. M., 2013, Solar forecasting methods for renewable energy integration, *Progress in Energy and Combustion Science*, 39 (6), 535-576.

IRENA, 2020, Renewable Energy Statistics 2020, *The International Renewable Energy Agency, Abu Dhabi*, 408.

Kadirgama, K., Amirruddin, A. K. ve Bakar, R. A., 2014, Estimation of Solar Radiation by Artificial Networks: East Coast Malaysia, *Energy Procedia*, 52, 383-388.

Karabacak, K., 2016, Rüzgâr-Fotovoltaik Hibrit Güç Sistemlerinin Yapay Sinir Ağları ile Kontrolü, Doktora Tezi, *Ege Üniversitesi*, İzmir, 105.

Kayri, M., Kayri, I. ve Gencoglu, M. T., 2017, The performance comparison of Multiple Linear Regression, Random Forest and Artificial Neural Network by using photovoltaic and atmospheric data, *2017 14th International Conference on Engineering of Modern Electric Systems (EMES)*, 1-4.

Khatib, T., Mohamed, A., Sopian, K. ve Mahmoud, M., 2012, Assessment of Artificial Neural Networks for Hourly Solar Radiation Prediction, *International Journal of Photoenergy*, 2012, 1-7.

Khosravi, A., Koury, R. N. N., Machado, L. ve Pabon, J. J. G., 2018a, Prediction of hourly solar radiation in Abu Musa Island using machine learning algorithms, *Journal of Cleaner Production*, 176, 63-75.

Khosravi, A., Nunes, R. O., Assad, M. E. H. ve Machado, L., 2018b, Comparison of artificial intelligence methods in estimation of daily global solar radiation, *Journal of Cleaner Production*, 194, 342-358.

Kim, S.-G., Jung, J.-Y. ve Sim, M., 2019, A Two-Step Approach to Solar Power Generation Prediction Based on Weather Data Using Machine Learning, *Sustainability*, 11 (5).

Ko, C.-N. ve Lee, C.-M., 2013, Short-term load forecasting using SVR (support vector regression)-based radial basis function neural network with dual extended Kalman filter, *Energy*, 49, 413-422.

Kumar, V. ve Minz, S., 2014, Feature Selection: A literature Review, *Smart Comput. Rev.*, 4, 211-229.

Lazzaroni, M., Ferrari, S., Piuri, V., Salman, A., Cristaldi, L. ve Faifer, M., 2015, Models for solar radiation prediction based on different measurement sites, *Measurement*, 63, 346-363.

Lazzús, J. A., Pérez Ponce, A. A. ve Marín, J., 2011, Estimation of global solar radiation over the city of La Serena (Chile) using a neural network, *Applied Solar Energy*, 47 (1), 66-73.

Li, D. H. W., Chen, W., Li, S. ve Lou, S., 2019, Estimation of hourly global solar radiation using Multivariate Adaptive Regression Spline (MARS) – A case study of Hong Kong, *Energy*, 186.

Long, H., Zhang, Z. ve Su, Y., 2014, Analysis of daily solar power prediction with data-driven approaches, *Applied Energy*, 126, 29-37.

Lotfinejad, M. M., Hafezi, R., Khanali, M., Hosseini, S. S., Mehrpooya, M. ve Shamshirband, S., 2018, A Comparative Assessment of Predicting Daily Solar Radiation Using Bat Neural Network (BNN), Generalized Regression Neural Network (GRNN), and Neuro-Fuzzy (NF) System: A Case Study, *Energies*, 11 (5).

Loutfi, H., Bernatchou, A. ve Tadili, R., 2017a, Generation of Horizontal Hourly Global Solar Radiation From Exogenous Variables Using an Artificial Neural Network in Fes (Morocco), *International Journal of Renewable Energy Research-IJRER*, 7 (3), 1097-1107.

Loutfi, H., Bernatchou, A. ve Tadili, R., 2017b, Generation of horizontal hourly global solar radiation from exogenous variables using an Artificial Neural Network in Fes (Morocco), *International Journal of Renewable Energy Research*, 7, 1097-1107.

Makridakis, S., Spiliotis, E. ve Assimakopoulos, V., 2018, Statistical and Machine Learning forecasting methods: Concerns and ways forward, *PLoS One*, 13 (3), e0194889.

MathWorks, 2018, What Is Machine Learning?, https://www.mathworks.com/discovery/machine-learning.html: [09.01.2021].

Meenal, R. ve Selvakumar, A. I., 2018, Assessment of SVM, empirical and ANN based solar radiation prediction models with most influencing input parameters, *Renewable Energy*, 121, 324-343.

MGM, 2020, İllere Ait Mevsim Normalleri (1991-2020), Meteoroloji Genel Müdürlüğü-Resmi İstatistikler Sayfası, https://www.mgm.gov.tr/veridegerlendirme/il-ve-ilceler-istatistik.aspx: [19.05.2020].

Moghaddamnia, A., Remesan, R., Kashani, M. H., Mohammadi, M., Han, D. ve Piri, J., 2009, Comparison of LLR, MLP, Elman, NNARX and ANFIS Models—with a case study in solar radiation estimation, *Journal of Atmospheric and Solar-Terrestrial Physics*, 71 (8-9), 975-982.

Mohandes, M., Rehman, S. ve Halawani, T. O., 1998, Estimation of global solar radiation using artificial neural networks, *Renewable Energy*, 14 (1), 179-184.

Mohandes, M., Balghonaim, A., Kassas, M., Rehman, S. ve Halawani, T. O., 2000, Use of radial basis functions for estimating monthly mean daily solar radiation, *Solar Energy*, 68 (2), 161-168.

Mori, H. ve Takahashi, A., 2012, A data mining method for selecting input variables for forecasting model of global solar radiation, *PES T&D 2012*, 1-6.

Mosavi, A., Salimi, M., Faizollahzadeh Ardabili, S., Rabczuk, T., Shamshirband, S. ve Varkonyi-Koczy, A., 2019, State of the Art of Machine Learning Models in Energy Systems, a Systematic Review, *Energies*, 12 (7).

Mubiru, J. ve Banda, E. J. K. B., 2008, Estimation of monthly average daily global solar irradiation using artificial neural networks, *Solar Energy*, 82 (2), 181-187.

Muneer, T., Gago, E. J. ve Etxebarria, S., 2015, Monthly-averaged hourly solar diffuse radiation models for world-wide locations, *Future Cities and Environment*, 1 (0).

Notton, G., Paoli, C., Vasileva, S., Nivet, M. L., Canaletti, J.-L. ve Cristofari, C., 2012, Estimation of hourly global solar irradiation on tilted planes from horizontal one using artificial neural networks, *Energy*, 39 (1), 166-179.

Olatomiwa, L., Mekhilef, S., Shamshirband, S., Mohammadi, K., Petković, D. ve Sudheer, C., 2015, A support vector machine–firefly algorithm-based model for global solar radiation prediction, *Solar Energy*, 115, 632-644.

Ongsulee, P., 2017, Artificial intelligence, machine learning and deep learning, *2017 15th International Conference on ICT and Knowledge Engineering (ICT&KE)*, 1-6.

Ozgoren, M., Bilgili, M. ve Sahin, B., 2012, Estimation of global solar radiation using ANN over Turkey, *Expert Systems with Applications*, 39 (5), 5043-5051.

Ozoegwu, C. G., 2019, Artificial neural network forecast of monthly mean daily global solar radiation of selected locations based on time series and month number, *Journal of Cleaner Production*, 216, 1-13.

Pang, Z., Niu, F. ve O'Neill, Z., 2020, Solar radiation prediction using recurrent neural network and artificial neural network: A case study with comparisons, *Renewable Energy*, 156, 279-289.

Paoli, C., Voyant, C., Muselli, M. ve Nivet, M.-L., 2010, Forecasting of preprocessed daily solar radiation time series using neural networks, *Solar Energy*, 84 (12), 2146-2160.

Perez, P. F. J., 2016, Data Mining Models For Short-Term Solar Radiation Prediction and Forecast-Based Assessment of Photovoltaic Facilities, Doktora Tezi, *Malaga Üniversitesi*, Malaga, İspamya, 178.

Piri, J., Shamshirband, S., Petković, D., Tong, C. W. ve Rehman, M. H. u., 2015, Prediction of the solar radiation on the Earth using support vector regression technique, *Infrared Physics & Technology*, 68, 179-185.

Premalatha, N. ve Valan Arasu, A., 2016, Prediction of solar radiation for solar systems by using ANN models with different back propagation algorithms, *Journal of Applied Research and Technology*, 14 (3), 206-214.

Prescott, J. A., 1940, Evaporation from water surface in relation to solar radiation, *Transactions of the Royal Society of Australia*, 46, 114-118.

Qazi, A., Fayaz, H., Wadi, A., Raj, R. G., Rahim, N. A. ve Khan, W. A., 2015, The artificial neural network for solar radiation prediction and designing solar systems: a systematic literature review, *Journal of Cleaner Production*, 104, 1-12.

Quinlan, J. R., 1992, Learning with continuous classes, *5th Australian Joint Conference on Artificial Intelligence (AI '92)*, Singapore: World Scientific, 343-348.

Raja, D., 2019, Solar Radiation Measurement Methods using Pyrheliometer and Pyranometer, CircuitDigiest, https://circuitdigest.com/tutorial/solar-radiation-measurement-methods-using-pyrheliometer-and-pyranometer: [07.02.2021].

Rao K, D. V. S. K., Premalatha, M. ve Naveen, C., 2018, Analysis of different combinations of meteorological parameters in predicting the horizontal global solar radiation with ANN approach: A case study, *Renewable and Sustainable Energy Reviews*, 91, 248-258.

Reed, R. D. ve Marks, R. J., 1999, Neural Smithing: Supervised Learning in Feedforward Artificial Neural Networks, *Cambridge, United States*, MIT Press, p. 360.

REN21, 2020, Renewables 2020 Global Status Report, *Paris: REN21 Secretariat, France*, 367.

Sağıroğlu, Ç., 2016, Türkiye'nin İller Bazında Enerji İhtiyacının Veri Madenciliği ile Analizi ve Lisanssız Yenilenebilir Enerji Yatırımları için Karar Destek Önermesi, Doktora Tezi, *Çukurova Üniveristesi*, Adana, 201.

Salcedo-Sanz, S., Casanova-Mateo, C., Pastor-Sánchez, A. ve Sánchez-Girón, M., 2014, Daily global solar radiation prediction based on a hybrid Coral Reefs Optimization – Extreme Learning Machine approach, *Solar Energy*, 105, 91-98.

Sammut, C. ve Webb, G. I., 2010, Encyclopedia of Machine Learning, *Boston*, Springer Publishing Company, Incorporated, p. 1058.

Selbaş, R., Şencan, A. ve Küçüksille, E. U., 2011, Data Mining Method For Energy System Aplications, In: Knowledge-Oriented Applications in Data Mining, Eds: Funatsu, K., *Rijeka*: IntechOpen, p. 339-352.

Sevim, S., 2019, Makine Öğrenmesi Yöntemiyle Hastalıkların Önceden Belirlenmesi ve Diyabet Üzerine Bir Uygulama, Yüksek Lisans Tezi, *İstanbul Aydın Üniversitesi*, İstanbul, 78.

Shadab, A., Said, S. ve Ahmad, S., 2019, Box–Jenkins multiplicative ARIMA modeling for prediction of solar radiation: a case study, *International Journal of Energy and Water Resources*, 3 (4), 305-318.

Singh, G. K., 2013, Solar power generation by PV (photovoltaic) technology: A review, *Energy*, 53, 1-13.

Sobri, S., Koohi-Kamali, S. ve Rahim, N. A., 2018, Solar photovoltaic generation forecasting methods: A review, *Energy Conversion and Management*, 156, 459-497.

Sözen, A., Arcaklioğlu, E., Özalp, M. ve Kanit, E. G., 2004, Use of artificial neural networks for mapping of solar potential in Turkey, *Applied Energy*, 77 (3), 273-286.

Sözen, A., Arcaklıoğlu, E., Özalp, M. ve Çağlar, N., 2005, Forecasting based on neural network approach of solar potential in Turkey, *Renewable Energy*, 30 (7), 1075-1090.

Susanto, A. ve Meiryani, 2019, Functions, Processes, Stages And Application Of Data Mining, *INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH*, 8 (07), 136-140.

Şahin, M., 2019, Sivas İlinin Güneş Işınım Şiddeti ve Rüzgâr Hızının Yapay Sinir Ağları ile Tahmin Edilmesi, Yüksek Lisans Tezi, *Cumhuriyet Üniversitesi*, Sivas, 65.

Şahingil, M. C., 2019, Veri Madenciliği Teknikleri Kullanılarak Gen Regülasyonunun İncelenmesi, Doktora Tezi, *Hacettepe Üniversitesi*, Ankara, 351.

Tasdemir, S., Yaniktepe, B. ve Guher, A. B., 2018, The effect on the wind power performance of different normalization methods by using multilayer feed-forward backpropagation neural network, 3, p.

Taşdemir, Ş., Yanıktepe, B. ve Güher, A. B., 2017, Determination of wind potential of a specific region using artificial neural networks.

TMMOB, 2018, Türkiye'nin Enerji Görünümü 2018 Raporu, *Makine Mühendisleri Odası*, *Ankara*, 614.

TMMOB, 2020, Türkiye'nin Enerji Görünümü 2020 Raporu, *Makina Mühendisleri Odası*, *Ankara*, 482.

Turkey_Investment_Office, 2019, Energy Industry Report, Ankara, https://www.invest.gov.tr/en/library/publications/lists/investpublications/energy-industry.pdf: [02.11].

Uyan, M., 2017, Optimal site selection for solar power plants using multi-criteria evaluation: A case study from the Ayranci region in Karaman, Turkey, *Clean Technologies and Environmental Policy*, 19 (9), 2231-2244.

Vapnik, V. N., 2000, The Nature of Statistical Learning Theory. New York, NY, Springer**:** 314.

Venkatesh, B. ve Anuradha, J., 2019, A Review of Feature Selection and Its Methods, *Cybernetics and Information Technologies*, 19 (1), 3-26.

Voyant, C., Notton, G., Kalogirou, S., Nivet, M.-L., Paoli, C., Motte, F. ve Fouilloy, A., 2017, Machine learning methods for solar radiation forecasting: A review, *Renewable Energy*, 105, 569-582.

Wang, L., Kisi, O., Zounemat-Kermani, M., Salazar, G. A., Zhu, Z. ve Gong, W., 2016, Solar radiation prediction using different techniques: model evaluation and comparison, *Renewable and Sustainable Energy Reviews*, 61, 384-397.

Wang, L., Kisi, O., Zounemat-Kermani, M., Zhu, Z., Gong, W., Niu, Z., Liu, H. ve Liu, Z., 2017, Prediction of solar radiation in China using different adaptive neuro-fuzzy methods and M5 model tree, *International Journal of Climatology*, 37 (3), 1141-1155.

Wang, Z., Wang, F. ve Su, S., 2011, Solar Irradiance Short-Term Prediction Model Based on BP Neural Network, *Energy Procedia*, 12, 488-494.

WideSkills, 2015, 04 - Data Mining Processes, WideSkills-Data Mining Tutorial, https://www.wideskills.com/data-mining-tutorial/data-mining-processes: [22.05.2019].

Witten, I. H., Frank, E., Trigg, L. E., Hall, M. A., Holmes, G. ve Cunningham, S. J., 1999, Weka: Practical machine learning tools and techniques with Java implementations.

Witten, I. H. ve Frank, E., 2005, Data Mining: Practical Machine Learning Tools and Techniques (Second Edition), *Boston*, Morgan Kaufmann, p. 560.

Witten, I. H., Frank, E. ve Hall, M. A., 2011, Chapter 1 - What's It All About?, In: Data Mining: Practical Machine Learning Tools and Techniques (Third Edition), Eds: Witten, I. H., Frank, E. ve Hall, M. A., *Boston*: Morgan Kaufmann, p. 3-38.

Yadav, A. K., Malik, H. ve Chandel, S. S., 2014, Selection of most relevant input parameters using WEKA for artificial neural network based solar radiation prediction models, *Renewable and Sustainable Energy Reviews*, 31, 509-519.

Yadav, A. K. ve Chandel, S. S., 2015, Solar energy potential assessment of western Himalayan Indian state of Himachal Pradesh using J48 algorithm of WEKA in ANN based prediction model, *Renewable Energy*, 75, 675-693.

Yadav, A. K., Malik, H. ve Chandel, S. S., 2015, Application of rapid miner in ANN based prediction of solar radiation for assessment of solar energy resource potential of 76 sites in Northwestern India, *Renewable and Sustainable Energy Reviews*, 52, 1093-1106.

Yıldırım, H. B., 2016, Meteorolojik Veriler Kullanılarak Doğu Akdeniz Bölgesi'nde Güneş Işınımı Tahmin Modellerinin Oluşturulması, Doktora Tezi, *Çukurova Universitesi Fen Bilimleri Enstitüsü Elektrik Elektronik Mühendisliği*, 200.

Yıldırım, H. B., Çelik, Ö., Teke, A. ve Barutçu, B., 2018, Estimating daily Global solar radiation with graphical user interface in Eastern Mediterranean region of Turkey, *Renewable and Sustainable Energy Reviews*, 82, 1528-1537.

Yu, H. ve Wilamowski, B. M., 2011, Levenberg-marquardt training. Industrial electronics handbook. 5.

Yurtsever, F. G. ve Öztopal, A., 2017, Akdeniz Bölgesi'nin Güneş Enerjisi Potansiyelinin Uydu Verileriyle İncelenmesi. III. Meteorolojik Uzaktan Algılama Sempozyumu. UZMED. Antalya, Uzaktan Algılama Şube Müdürlüğü.

Zeng, J. ve Qiao, W., 2013, Short-term solar power prediction using a support vector machine, *Renewable Energy*, 52, 118-127.

Zhang, J., Zhao, L., Deng, S., Xu, W. ve Zhang, Y., 2017, A critical review of the models used to estimate solar radiation, *Renewable and Sustainable Energy Reviews*, 70, 314-329.

Zheng, A. ve Casari, A., 2018, Feature Engineering for Machine Learning: Principles and Techniques for Data Scientists, O'Reilly Media, Inc., p.

Zou, L., Wang, L., Lin, A., Zhu, H., Peng, Y. ve Zhao, Z., 2016, Estimation of global solar radiation using an artificial neural network based on an interpolation technique in southeast China, *Journal of Atmospheric and Solar-Terrestrial Physics*, 146, 110-122.

ÖZGEÇMİŞ

KİŞİSEL BİLGİLER

Adı Soyadı : Abdurrahman Burak GÜHER
Uyruğu : T.C.
Doğum Yeri ve Tarihi :
Telefon :
Faks :
E-mail :

EĞİTİM

Derece		**Adı, İlçe, İl**	**Bitirme Yılı**
Lise	:		
Üniversite	:		
Yüksek Lisans	:		
Doktora	:	Selçuk Üniversitesi, F.B.E. Bilgisayar Mühendisliği A.B.D.	2021

İŞ DENEYİMLERİ

Yıl	**Kurum**	**Görevi**

UZMANLIK ALANI

Veri madenciliği, Yapay Zekâ, Makine Öğrenme Algoritmaları, Derin Öğrenme.

YABANCI DİLLER

İngilizce (İyi Derecede)

YAYINLAR

Guher, A. B., Tasdemir, S. ve Yaniktepe, B., 2020, Effective Estimation of Hourly Global Solar Radiation Using Machine Learning Algorithms, *International Journal of Photoenergy*, 2020, 1-26. https://dx.doi.org/10.1155/2020/8843620 (SCI-E) (Doktora Tezinden Yapılmıştır)

Guher, A. B., Tasdemir, S. ve Yaniktepe, B., 2021, Determining of Solar Power by Using Machine Learning Methods in a Specified Region, *Tehnički vjesnik*, 28(5), https://doi.org/10.17559/TV-20200425151543. (SCI-E) (Basılacak) (Doktora Tezinden Yapılmıştır)

Tasdemir, S., Yaniktepe, B., ve **Guher, A. B**., 2018, The effect on the wind power performance of different normalization methods by using multilayer feed-forward backpropagation neural network, *International Journal of Energy Applications and Technologies*, 5(3) ,131-139. https://doi.org/ 10.31593/ijeat.464210

Tasdemir, S., Yaniktepe, B., ve **Guher, A. B**., 2017, Determination of Wind Potential of a Specific Region using Artificial Neural Networks, *International Journal of Intelligent Systems and Applications in Engineering*, 5(3), 158-162. https://doi.org/10.18201/ijisae.2017531433

Tasdemir, S., Yaniktepe, B., **Guher, A. B**., ve Akcan, S., 2016, Wind Power Forecasting For The Province Of Osmaniye Using Artificial Neural Network Method, *International Journal of Intelligent Systems and Applications in Engineering*, 4 (Special Issue), 114-117. https://doi.org/10.18201/ijisae.270560

Yanıktepe, B., Taşdemir, Ş., ve **Güher, A. B.**, 2017, Estimation of Short Term Measured Solar Radiation Potential by ANN Method, *III Internatıonal Conference On Engıneerıng And Natural Scıences (ICENS)*, 4-6 Mayıs 2017, Budapeşte. (Doktora Tezinden Yapılmıştır)

Yanıktepe, B., Taşdemir, Ş., ve **Güher, A. B.**, 2017, The Estimation of Wind Power with Feed Forward Artificial Neural Network, *8th International Advanced Technologies Symposium (IATS)*, 19-22 Ekim 2017, Elazığ.

Güher, A. B., 2017, E Öğrenme Platformları Kullanımı ve Uygulanmasının Bilgisayar ve Öğretim Teknolojileri Eğitimi Temelinde Değerlendirilmesi, *International World of Turks Symposium of Social Sciences*, 11-14 Mayıs 2017, Antalya.

Yanıktepe, B., Taşdemir, Ş., ve **Güher, A. B.**, 2017, Estimation of Short Period Wind Power with ANN Method, *5th International Conference on Advanced Technology Sciences*, 09-12 Mayıs 2017, İstanbul.

Yanıktepe, B., Taşdemir, Ş., ve **Güher, A. B.**, 2017, Evaluation of Estimation of Wind Power Potential in Backpropagation Networks in the Frame of Different Normalization Techniques, *8th International Advanced Technologies Symposium (IATS)*, 19-22 Ekim 2017, Elazığ.

Yanıktepe, B., Taşdemir, Ş., **Güher, A. B.**, ve Akcan, S., 2016, Wind Power Forecasting for The Province of Osmaniye Using Artificial Neural Network Method, *3rd International Conference on Advanced Technology & Sciences (ICAT'16)*, 1-3 Eylül 2016, Konya.